

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA QUẢN TRỊ KINH DOANH 1
BỘ MÔN KINH TẾ**

**BÀI GIẢNG
THỐNG KÊ DOANH NGHIỆP**

Người biên soạn : TS. Vũ Trọng Phong

Hà nội - 2017

MỤC LỤC

Trang

Lời nói đầu	9
Phần 1 Lý thuyết thống kê	10
Chương I: Các phương pháp trình bày số liệu thống kê	10
1.1. Đối tượng nghiên cứu của thống kê học	10
1.2. Một số khái niệm thống kê.....	10
1.2.1 Khái niệm thống kê.....	10
1.2.2 Tổng thể thống kê và đơn vị tổng thể thống kê.....	12
1.2.3 Tổng thể mẫu và quan sát.....	13
1.2.4 Tiêu thức thống kê (gọi tắt là tiêu thức).....	13
1.2.5 Chỉ tiêu thống kê (gọi tắt là chỉ tiêu).....	14
1.2.6 Thang đo trong thống kê.....	15
1.2.7 Hoạt động thống kê và quá trình nghiên cứu thống kê.....	16
1.3 Các phương pháp trình bày số liệu bằng thống kê.....	17
1.3.1 Bảng thống kê.....	17
1.3.2 Đồ thị thống kê.....	19
Tài liệu tham khảo chương 1.....	22
Câu hỏi ôn tập chương 1.....	22
Chương 2: Các mức độ của hiện tượng thống kê	23
2.1. Số tuyệt đối.....	23
2.1.1. Khái niệm và ý nghĩa của số tuyệt đối.....	23
2.1.2. Đặc điểm của số tuyệt đối.....	23
2.1.3. Phân loại số tuyệt đối.....	23
2.1.4 Đơn vị tính số tuyệt đối.....	24
2.2 Số tương đối	24
2.2.1. Khái niệm và ý nghĩa số tương đối.....	24
2.2.2. Đặc điểm của số tương đối.....	25

2.2.3. Các loại số tương đối.....	25
2.3. Số bình quân	27
2.3.1. Khái niệm số bình quân.....	27
2.3.2. Ý nghĩa của số bình quân.....	27
2.3.3. Các loại số bình quân trong thống kê.....	28
2.4. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức.....	37
2.4.1. Ý nghĩa nghiên cứu	37
2.4.2. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức.....	37
2.4.3 Các phương pháp tính phương sai.....	40
2.5. Phân phối trong thống kê.....	42
2.5.1. Một số phân phối lý thuyết.....	42
2.5.2. So sánh phân phối thực nghiệm với phân phối lý thuyết.....	45
2.5.3. Các chỉ tiêu biểu thị hình dáng của phân phối.....	47
Tài liệu tham khảo chương 2.....	49
Câu hỏi ôn tập chương 2.....	50
Chương 3: Điều tra chọn mẫu	53
3.1. Khái niệm, ưu nhược điểm và phạm vi sử dụng điều tra chọn mẫu.....	53
3.2. Điều tra chọn mẫu ngẫu nhiên.....	54
3.2.1 Tổng thể chung và tổng thể mẫu.....	54
3.2.2 Chọn mẫu với xác suất đều và xác suất không đều.....	56
3.2.3. Sai số trong chọn mẫu	56
3.2.4 Sai số bình quân chọn mẫu và phạm vi sai số chọn mẫu.....	57
3.3. Quy trình một cuộc điều tra chọn mẫu ngẫu nhiên.....	62
3.4. Điều tra chọn mẫu phi ngẫu nhiên.....	64
Tài liệu tham khảo chương 3.....	65
Câu hỏi ôn tập chương 3.....	65
Chương 4: Tương quan và hồi quy	67
4.1. Khái niệm	67

4.1.1. Tiêu thức nguyên nhân và tiêu thức kết quả.....	67
4.1.2. Nhiệm vụ phương pháp hồi quy và tương quan.....	67
4.2. Hồi quy tương quan tuyến tính giữa hai tiêu thức số lượng.....	68
4.3. Hồi quy tương quan phi tuyến giữa hai tiêu thức số lượng.....	72
4.3.1 Các dạng phương trình hồi qui.....	72
4.3.2. Các chỉ tiêu đánh giá mối liên hệ tương quan phi tuyến tính.....	73
4.4. Hồi quy tương quan tuyến tính bội.....	75
4.4.1 Mô hình tuyến tính bội.....	75
4.4.2 Đa cộng tuyến.....	79
Tài liệu tham khảo chương 4.....	80
Câu hỏi ôn tập chương 4.....	80
Chương 5: Dãy số thời gian	82
5.1. Khái niệm, phân loại và ý nghĩa của dãy số thời gian.....	82
5.1.1. Khái niệm.....	82
5.1.2. Phân loại dãy số thời gian.....	82
5.1.3. Ý nghĩa của dãy số thời gian.....	83
5.1.4. Yêu cầu đối với dãy số thời gian.....	83
5.2. Các chỉ tiêu phân tích dãy số thời gian.....	83
5.2.1 Mức độ bình quân theo thời gian.....	83
5.2.2. Lượng tăng (hoặc giảm) tuyệt đối.....	85
5.2.3.Tốc độ phát triển.....	87
5.2.4. Tốc độ tăng (hoặc giảm).....	89
5.2.5. Giá trị tuyệt đối của 1% tăng (hoặc giảm).....	91
5.3. Các phương pháp biểu hiện xu hướng phát triển của hiện tượng.....	91
5.3.1. Phương pháp mở rộng khoảng cách thời gian.....	91
5.3.2 Phương pháp số bình quân trượt (di động).....	92
5.3.3. Phương pháp hồi quy.....	93
5.3.4 Phương pháp biểu hiện biến động thời vụ.....	96

5.4. Dự báo thống kê ngắn hạn.....	97
5.4.1. Khái niệm và đặc điểm của dự báo thống kê.....	97
5.4.2. Các phương pháp dự báo thống kê.....	98
Tài liệu tham khảo chương 5.....	99
Câu hỏi ôn tập chương 5.....	99
Chương 6: Chỉ số trong thống kê	101
6.1. Khái niệm, ý nghĩa, phân loại chỉ số.....	101
6.1.1. Khái niệm chỉ số.....	101
6.1.2. Đặc điểm của chỉ số.....	101
6.1.3 Tác dụng của chỉ số.....	101
6.1.4. Phân loại chỉ số	102
6.2. Chỉ số phát triển.....	102
6.2.1. Chỉ số đơn.....	102
6.2.2. Chỉ số tổng hợp.....	103
6.2.3 Chỉ số không gian.....	107
6.3. Chỉ số kế hoạch.....	108
6.4 Hệ thống chỉ số.....	109
Tài liệu tham khảo chương 6.....	114
Câu hỏi ôn tập chương 6.....	114
Phần 2 Thống kê doanh nghiệp	115
Chương 7 Thống kê kết quả hoạt động kinh doanh của doanh nghiệp	115
7.1. Một số khái niệm cơ bản về kết quả hoạt động sản xuất kinh doanh.....	115
7.1.1 Kết quả sản xuất của doanh nghiệp.....	115
7.1.2. Các dạng biểu hiện kết quả hoạt động kinh doanh của một doanh nghiệp.....	115
7.1.3. Đơn vị đo lường kết quả hoạt động sản xuất kinh doanh của doanh nghiệp.....	116
7.2. Hệ thống chỉ tiêu đo lường kết quả hoạt động kinh doanh của doanh nghiệp.....	116
7.2.1. Giá trị sản xuất của doanh nghiệp (GO – Gross Output).....	117
7.2.2. Giá trị gia tăng của doanh nghiệp (VA – Value Added).....	119

7.2.3 Chi phí trung gian của doanh nghiệp (IC – Intermediational cost).....	120
7.2.4. Giá trị gia tăng thuần của doanh nghiệp (NVA – Net Value Added).....	122
7.2.5 Lợi nhuận (hay lãi) kinh doanh của doanh nghiệp (M).....	123
7.2.6 Doanh thu bán hàng	124
7.3. Phương pháp phân tích thống kê kết quả hoạt động sản xuất của doanh nghiệp	125
7.3.1 Phân tích kết cấu kết quả sản xuất của doanh nghiệp.....	125
7.3.2. Phân tích sự biến động theo thời gian của các chỉ tiêu phản ánh kết quả sản xuất kinh doanh của doanh nghiệp.....	128
Tài liệu tham khảo chương 7.....	126
Câu hỏi ôn tập chương 7.....	127
Chương 8 Thống kê lao động của doanh nghiệp	128
8.1 Thống kê số lượng và sự biến động lao động của doanh nghiệp.....	128
8.1.1. Thống kê số lượng lao động của doanh nghiệp.....	128
8.1.2 Thống kê kết cấu lao động	131
8.1.3 Thống kê nghiên cứu biến động số lượng lao động.....	132
8.2. Thống kê tình hình sử dụng số lượng và thời gian lao động của doanh nghiệp.....	133
8.2.1. Thống kê tình hình sử dụng số lượng lao động của doanh nghiệp.....	133
8.2.2 Thống kê sử dụng thời gian lao động.....	134
8.3. Thống kê năng suất lao động.....	139
8.3.1 Năng suất lao động và nhiệm vụ thống kê.....	139
8.3.2 Thống kê tính toán chỉ tiêu năng suất lao động.....	139
8.3.3 Thống kê nghiên cứu biến động năng suất lao động.....	140
Tài liệu tham khảo chương 8.....	141
Câu hỏi ôn tập chương 8.....	141
Chương 9 Thống kê tài sản doanh nghiệp	143
9.1. Thống kê tài sản cố định.....	143
9.1.1 Khái niệm tài sản cố định	143
9.1.2. Phân loại tài sản cố định	143
9.1.3. Đánh giá tài sản cố định	145

9.1.4. Thống kê số lượng tài sản cố định	146
9.1.5 Thống kê kết cấu TSCĐ.....	147
9.1.6 Thống kê biến động TSCĐ.....	148
9.1.7 Thống kê trạng thái TSCĐ.....	149
9.1.8 Thống kê tình hình trang bị, sử dụng và hiệu quả sử dụng TSCĐ	150
9.2 Thống kê khấu hao tài sản cố định	151
9.2.1 Một số khái niệm	151
9.2.2 Phương pháp khấu hao tài sản cố định.....	151
Tài liệu tham khảo chương 9.....	152
Câu hỏi ôn tập chương 9.....	152
Chương 10 Thống kê giá thành của doanh nghiệp	154
10.1. Khái niệm, ý nghĩa của các loại chỉ tiêu giá thành.....	154
10.1.1 Khái niệm, ý nghĩa của chỉ tiêu giá thành tổng hợp.....	154
10.1.2. Các loại chỉ tiêu giá thành và ý nghĩa của nó đối với công tác quản lý doanh nghiệp.....	154
10.2. Nội dung kinh tế của chỉ tiêu giá thành.....	156
10.2.1. Xét về nội dung kinh tế của chỉ tiêu giá thành.....	156
10.2.2 Xét chi phí theo công dụng cụ thể của chi phí trong sản xuất.....	157
10.3. Phương pháp phân tích tài liệu thống kê giá thành.....	157
10.3.1. Phân tích cấu thành của chỉ tiêu giá thành.....	157
10.3.2. Phân tích sự biến động cấu thành của chỉ tiêu giá thành theo thời gian.....	158
10.3.3. Phân tích nhân tố ảnh hưởng đến giá thành bằng phương pháp hồi quy và tương quan.....	158
10.3.4. Phân tích nhân tố ảnh hưởng đến giá thành bình quân	158
10.3.5 Nghiên cứu biến động giá thành sản phẩm dịch vụ theo thời gian.....	159
Tài liệu tham khảo chương 10.....	160
Câu hỏi ôn tập chương 10.....	160
Chương 11 Thống kê vốn và hoạt động tài chính của doanh nghiệp	161
11.1. Thông kê vốn đầu tư của doanh nghiệp.....	161

11.1.1 Khái niệm về đầu tư và vốn đầu tư cơ bản của doanh nghiệp.....	161
11.1.2. Thống kê khối lượng vốn đầu tư xây dựng cơ bản.....	162
11.1.3 Thống kê biến động khối lượng vốn đầu tư cơ bản	164
11.2. Thống kê vốn sản xuất kinh doanh của doanh nghiệp	165
11.2.1 Thống kê vốn cố định	165
11.2.2. Thống kê vốn lưu động.....	169
11.3. Thống kê kết quả hoạt động tài chính của doanh nghiệp.....	174
11.3.1. Thống kê mức độ độc lập về tài chính của doanh nghiệp.....	175
11.3.2 Thống kê khả năng thanh toán công nợ và tình hình chiếm dụng vốn của doanh nghiệp.....	175
Tài liệu tham khảo chương 11.....	177
Câu hỏi ôn tập chương 11.....	177

LỜI MỞ ĐẦU

Trong điều kiện hoạt động kinh doanh theo cơ chế thị trường, muốn tồn tại và phát triển, đòi hỏi doanh nghiệp nói chung, doanh nghiệp bưu chính viễn thông nói riêng phải có đầy đủ thông tin. Muốn vậy, doanh nghiệp phải tổ chức thống kê, thu thập thông tin.

*Thống kê doanh nghiệp là một môn học không thể thiếu được trong chương trình đào tạo khối ngành kinh tế và quản trị kinh doanh. Để đáp ứng yêu cầu giảng dạy và học tập của giáo viên và sinh viên, chúng tôi tổ chức biên soạn bài giảng "**Thống kê doanh nghiệp**". Với kinh nghiệm giảng, cộng với sự nỗ lực nghiên cứu từ các nguồn tài liệu khác nhau, bài giảng có nhiều thay đổi bổ sung để đáp ứng yêu cầu đào tạo trong giai đoạn mới.*

*Bài giảng gồm 11 chương đề cập đến toàn bộ những kiến thức về nguyên lý thống kê và Thống kê doanh nghiệp. Bài giảng được hiệu chỉnh dựa trên bài giảng "**Thống kê doanh nghiệp**" đã được biên soạn năm 2013 có chỉnh lý và bổ sung thêm các nội dung câu hỏi và bài tập cho các chương.*

Tác giả chân thành cảm ơn các đồng nghiệp đã giúp đỡ trong quá trình biên soạn bài giảng này. Tác giả mong muốn nhận được góp ý của đồng nghiệp, anh chị em sinh viên và bạn đọc để tiếp tục hoàn thiện nhằm nâng cao hơn nữa chất lượng bài giảng.

Hà Nội tháng 11 năm 2017

TÁC GIẢ

PHẦN I – LÝ THUYẾT THỐNG KÊ

CHƯƠNG 1

NHỮNG VẤN ĐỀ CƠ BẢN CỦA THỐNG KÊ HỌC

1.1. Đối tượng nghiên cứu của thống kê học

Đối tượng nghiên cứu của thống kê học là mặt lượng trong sự xác định về mặt chất của các hiện tượng và quá trình kinh tế xã hội số lớn, nghiên cứu cấu trúc, sự phân bố và vị trí của chúng trong không gian, sự biến động theo thời gian để chỉ ra bản chất và tính quy luật vốn có của chúng trong điều kiện thời gian và không gian cụ thể.

1.2. Một số khái niệm thống kê

1.2.1 Khái niệm thống kê

Nghiên cứu quá trình hình thành và phát triển của thống kê cho thấy: Thống kê học là một môn khoa học xã hội. Tuy nhiên, khác với các môn học xã hội khác, thống kê học không trực tiếp nghiên cứu mặt chất của hiện tượng mà nó chỉ phản ánh bản chất, tính quy luật của hiện tượng thông qua các con số, các biểu hiện về lượng của hiện tượng. Điều đó có nghĩa là thống kê học phải sử dụng các con số về quy mô, kết cấu quan hệ tỉ lệ, quan hệ so sánh, trình độ phát triển, trình độ phổ biến,... của hiện tượng để phản ánh, biểu thị bản chất, tính quy luật của hiện tượng nghiên cứu trong những điều kiện, hoàn cảnh cụ thể. Như vậy, các con số thống kê không phải chung chung, trừu tượng mà bao giờ cũng chứa đựng một nội dung kinh tế, chính trị, xã hội nhất định, giúp chúng ta nhận thức được bản chất và quy luật của hiện tượng nghiên cứu.

Theo quan điểm triết học, chất và lượng là hai mặt không thể tách rời của mọi sự vật, hiện tượng, giữa chúng luôn tồn tại mối liên hệ biện chứng với nhau. Trong mỗi quan hệ đó, sự thay đổi về lượng quyết định sự biến đổi về chất. Quy luật lượng - chất của triết học đã chỉ rõ: mỗi lượng cụ thể đều gắn với một chất nhất định, khi lượng thay đổi và tích lũy đến một chừng mực nhất định thì chất thay đổi theo. Vì vậy, nghiên cứu mặt lượng của hiện tượng sẽ giúp cho việc nhận thức bản chất của hiện tượng. Có thể đánh giá thành tích sản xuất của một doanh nghiệp qua các con số thống kê về tổng số sản phẩm làm ra, giá trị sản xuất đạt được, tỉ lệ hoàn thành kế hoạch sản xuất, giá thành đơn vị sản phẩm, năng suất lao động và thu nhập của công nhân.

Tuy nhiên, để có thể phản ánh được bản chất và quy luật phát triển của hiện tượng, các con số thống kê phải được tập hợp, thu thập trên một số lớn các hiện tượng cá biệt. Thống kê học coi tổng thể các hiện tượng cá biệt như một thể hoàn chỉnh và lấy đó làm đối tượng nghiên cứu. Mặt lượng của hiện tượng cá biệt thường chịu tác động của nhiều nhân tố, trong đó có những nhân tố tất nhiên và ngẫu nhiên. Mức độ và chiều hướng tác động của từng nhân tố này trên mỗi hiện tượng cá biệt là rất khác nhau. Nếu chỉ thu thập số liệu trên một số ít hiện tượng thì khó có thể rút ra bản chất chung của hiện tượng, mà nhiều khi người ta chỉ tìm thấy những yếu tố ngẫu nhiên, không bản chất. Ngược lại, khi nghiên cứu trên một số lớn các hiện tượng cá biệt, các yếu tố ngẫu nhiên sẽ bù trừ, triệt tiêu nhau và khi đó, bản chất, quy luật phát triển của hiện tượng mới được bộc lộ rõ.

Hiện tượng số lớn trong thống kê được hiểu là một tập hợp các hiện tượng cá biệt đủ bù trừ, triệt tiêu tác động của các yếu tố ngẫu nhiên. Giữa hiện tượng số lớn (tổng thể) và các hiện tượng cá biệt (đơn vị tổng thể) luôn tồn tại mối quan hệ biện chứng. Muốn nghiên cứu tổng thể, phải dựa trên cơ sở nghiên cứu từng đơn vị tổng thể. Mặt khác, trong quá trình phát triển không ngừng của xã hội, luôn nảy sinh những hiện tượng cá biệt mới, những điển hình tiên tiến hoặc lạc hậu. Sự nghiên cứu các hiện tượng cá biệt này sẽ giúp cho sự nhận thức bản chất của hiện tượng đầy đủ, toàn diện và sâu sắc hơn. Vì vậy trong thống kê, người ta thường kết hợp nghiên cứu hiện tượng số lớn với việc nghiên cứu hiện tượng cá biệt.

Như vậy, thuật ngữ thống kê có thể hiểu theo hai nghĩa:

- **Thứ nhất**, thống kê là các số liệu được thu thập để phản ánh các hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội. Chẳng hạn như sản lượng sản phẩm được sản xuất ra của một doanh nghiệp nào đó, mức nước trên một dòng sông ở các tháng trong năm,...

- **Thứ hai**, thống kê được hiểu là hệ thống các phương pháp ghi chép, thu thập và phân tích các con số về những hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội để tìm hiểu bản chất và tìm quy luật vốn có của những hiện tượng ấy.

Công việc của một nhà thống kê bao gồm các hoạt động trên một phạm vi rộng có thể tóm tắt thành những mục lớn như sau:

- Thu thập và xử lý dữ liệu.
- Điều tra chọn mẫu.
- Nghiên cứu mối liên hệ giữa các hiện tượng.
- Dự đoán.
- Nghiên cứu các hiện tượng trong các hoàn cảnh không chắc chắn
- Ra quyết định trong điều kiện không chắc chắn

Một cách tổng quát thống kê học chính là *khoa học nghiên cứu hệ thống các phương pháp thu thập, xử lý và phân tích các con số (mặt lượng) của những hiện tượng số lớn để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể.*

Hiện tượng bao giờ cũng có hai mặt chất và lượng không tách rời nhau. Chất của hiện tượng giúp ta phân biệt hiện tượng này với hiện tượng khác, đồng thời bộc lộ những khía cạnh sâu kín của hiện tượng. Nhưng chất không tồn tại độc lập mà được biểu hiện qua lượng, với những cách xử lý mặt lượng đó một cách khoa học. Sở dĩ cần phải xử lý mặt lượng mới tìm hiểu được mặt chất là vì mặt chất của hiện tượng thường bị che khuất dưới các tác động ngẫu nhiên. Phải thông qua tổng hợp mặt lượng của số lớn đơn vị cấu thành hiện tượng, tác động của các yếu tố ngẫu nhiên mới được bù trừ và triệt tiêu. Hơn nữa, cũng còn phải sử dụng các phương pháp phân tích số liệu thích hợp, bản chất của hiện tượng mới dần dần bộc lộ qua tính quy luật thống kê. Về thực chất, tính quy luật thống kê là sự biểu hiện về lượng của các quy luật phát sinh, phát triển của hiện tượng. Tính quy luật này không có tính chất chung chung mà rất cụ thể theo các điều kiện, địa điểm và thời gian cụ thể. Đó chính là đặc trưng của thống kê học, làm cho nó khác

với toán học. Tính quy luật thống kê có ý nghĩa rất quan trọng đối hoạt động kinh doanh, vì nó cho biết mối liên hệ giữa các hiện tượng, xu thế phát triển của hiện tượng cũng như các dao động chu kỳ của hiện tượng đó, quy luật phân phối của các tổng thể chứa đựng hiện tượng đang nghiên cứu.

Thống kê được chia thành hai lĩnh vực:

- Thống kê mô tả: bao gồm các phương pháp thu thập số liệu, mô tả và trình bày số liệu, tính toán các đặc trưng đo lường.

- Thống kê suy diễn: bao gồm các phương pháp ước lượng, kiểm định phân tích mối quan hệ, dự đoán trên cơ sở các thông tin thu thập từ mẫu.

1.2.2. Tổng thể thống kê và đơn vị tổng thể thống kê

Thống kê nghiên cứu một lượng của hiện tượng kinh tế xã hội số lớn phải xác định phạm vi hiện tượng được nghiên cứu cụ thể. Để chỉ đối tượng nghiên cứu cụ thể, người ta dùng khái niệm tổng thể. Tổng thể thống kê là hiện tượng kinh tế xã hội số lớn, gồm những đơn vị (hoặc phân tử, hiện tượng) cá biệt cần được quan sát, phân tích mặt lượng của chúng để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể.

Có trường hợp các đơn vị cấu thành tổng thể, có thể thấy được bằng trực quan. Tổng thể bao gồm các đơn vị như vậy được gọi là tổng thể bộc lộ như các tổng thể nêu trên. Tổng thể các đơn vị cấu thành nó, không thể nhận biết được bằng trực quan là tổng thể tiềm ẩn như tổng thể những cán bộ công nhân viên ưa chuộng nghệ thuật sân khấu, tổng thể những người mê tín dị đoan, tổng thể những người trung thành với Tổ quốc...

Các đơn vị tổng thể có thể giống nhau trên một số đặc điểm, các đặc điểm còn lại khác nhau. Do đó, tùy theo mục đích nghiên cứu mà phân biệt tổng thể đồng chất hay không đồng chất. Tổng thể đồng chất bao gồm các đơn vị giống nhau về một số đặc điểm chủ yếu có liên quan tới mục đích nghiên cứu tổng thể không đồng chất bao gồm các đơn vị khác nhau về các đặc điểm, các loại hình. Tổng thể bao gồm tất cả các đơn vị thuộc phạm vi nghiên cứu tổng thể chung, chỉ bao gồm một bộ phận đơn vị trong đó là tổng thể bộ phận.

Xác định tổng thể để đáp ứng mục đích nghiên cứu thống kê. Phải trên cơ sở phân tích lý luận kinh tế, chính trị hoặc xã hội, định nghĩa rõ tổng thể. Định nghĩa tổng thể không những phải giới hạn về thực thể (tổng thể là gì) mà còn phải giới hạn về thời gian và không gian (tổng thể tồn tại vào thời gian nào, ở đâu). Xác định tổng thể chính xác không dễ dàng. Vì có những hiện tượng có thể tương tự về hình thức, nhưng lại khác hẳn về nội dung. Chính vì vậy phải phân tích lý luận để thấy rõ nội dung của hiện tượng. Xác định tổng thể thống kê không chính xác sẽ lãng phí sức người và tiền của trong nghiên cứu, không đủ cơ sở để hiểu đúng bản chất cụ thể của hiện tượng.

Đúng nghĩa tổng thể làm rõ đặc trưng cơ bản chung của hiện tượng kinh tế xã hội, số lớn phù hợp với mục đích nghiên cứu. Thông qua việc phân tích lý luận và thực tế phải làm rõ tổng thể gồm những hiện tượng (phần tử) cá biệt nào. Hiện tượng cá biệt này là đơn vị tổng thể. Tất cả

các đơn vị tổng thể chỉ giống nhau trên một số mặt, còn các mặt khác không giống nhau. Cho nên trong thực tế phải nêu rõ ràng những hiện tượng cá biệt nào được kể là đơn vị tổng thể. Trong những trường hợp khó khăn cho việc giới hạn, người ta phải lập một danh mục các đơn vị hoặc trong giải thích cần xác định rõ phạm vi nào của các đơn vị thuộc tổng thể.

Đơn vị tổng thể bao giờ cũng có đơn vị tính toán phù hợp. Xác định đơn vị tổng thể là việc cụ thể hóa tổng thể. Đơn vị tổng thể là xuất phát điểm của quá trình nghiên cứu thống kê. Vì nó có mặt lượng mà ta cần nghiên cứu. Cho nên xác định đơn vị tổng thể cũng quan trọng như xác định tổng thể.

1.2.3. Tổng thể mẫu và quan sát

Tổng thể mẫu là tổng thể bao gồm một số đơn vị được chọn ra từ tổng thể chung theo một phương pháp lấy mẫu nào đó. Các đặc trưng mẫu được suy rộng ra các đặc trưng của tổng thể chung.

Quan sát là cơ sở để thu thập số liệu và thông tin cần nghiên cứu. Chẳng hạn trong điều tra chọn mẫu, mỗi đơn vị mẫu sẽ được tiến hành ghi chép, thu thập thông tin và được gọi là một quan sát.

1.2.4. Tiêu thức thống kê (gọi tắt là tiêu thức)

Nghiên cứu thống kê phải dựa vào các đặc điểm của đơn vị tổng thể. Đơn vị tổng thể có nhiều đặc điểm. Tùy theo mục đích nghiên cứu, một số đặc điểm của đơn vị tổng thể được chọn ra để nghiên cứu. Các đặc điểm này được gọi là các tiêu thức. Ví dụ mỗi cán bộ công nhân viên của bưu cục có các tiêu thức: tên, tuổi, giới tính, trình độ văn hóa, nghề nghiệp, nơi ở ... Mỗi bưu cục trong tổng thể có tiêu thức: tên bưu cục, địa chỉ, số lượng cán bộ công nhân viên... Đơn vị tổng thể được làm rõ đặc trưng của nó qua các tiêu thức: thực thể, thời gian và không gian.

1. Tiêu thức thực thể

Nêu lên bản chất của đơn vị tổng thể. Nó biến đổi trong bản chất này. Các tiêu thức: giới tính, tuổi, trình độ văn hóa, nghề nghiệp, số lượng cán bộ công nhân viên chức là các tiêu thức thực thể. Theo nội dung của nó, tiêu thức thực thể gồm hai loại: thuộc tính và số lượng.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là các con số, như tiêu thức giới tính, trình độ văn hóa, nghề nghiệp ... tiêu thức thuộc tính có biểu hiện trực tiếp và gián tiếp, như giới tính có biểu hiện trực tiếp là nam và nữ. Tiêu thức đời sống vật chất có biểu hiện gián tiếp là lương tiêu dùng lương thực, thực phẩm theo đầu người, diện tích nhà ở theo đầu người. Các biểu hiện gián tiếp của tiêu thức thuộc tính còn được gọi là các chỉ báo thống kê.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là con số, nên còn được gọi là tiêu thức phi lượng hóa.

Tiêu thức số lượng có biểu hiện trực tiếp là con số (gọi là lượng biến). Nó là kết quả của quá trình quan sát (cân đo, đong đếm) như tuổi đời, tuổi nghề, số lượng điện thoại, số bưu cục ... Tiêu thức số lượng còn gọi là tiêu thức lượng hóa vì nó có biểu hiện trực tiếp là con số.

Tiêu thức thực thể khi chỉ có hai biểu hiện không trùng nhau trên một đơn vị tổng thể, được gọi là tiêu thức thay phiên, như giới tính (nam và nữ), chất lượng sản phẩm dịch vụ (tốt và xấu). Tiêu thức thực thể có ba loại biểu hiện trở lên có thể trở thành tiêu thức thay phiên, như số lượng cán bộ công nhân viên nêu trên có nhiều biểu hiện nhưng rút gọn thành hai biểu hiện, hoặc các biểu hiện của tiêu thức trình độ văn hóa có thể rút gọn thành hai biểu hiện: chưa tốt nghiệp phổ thông trung học và tốt nghiệp phổ thông trung học trở lên. Những trường hợp này được tiến hành khi người ta chỉ quan tâm đến một biểu hiện nào đó xuất hiện hay không xuất hiện trên đơn vị tổng thể. Tiêu thức thực thể có phù hợp nhiều hay ít với việc đáp ứng mục đích nghiên cứu là tùy thuộc vào việc chọn những tiêu thức nào cho nghiên cứu.

2. Tiêu thức thời gian

Nêu hiện tượng kinh tế xã hội theo sự xuất hiện của nó vào thời gian nào. Những biểu hiện của tiêu thức thời gian là phút, giờ, ngày, tháng, năm. Thời hạn có giá trị của các chỉ dẫn về đối tượng nghiên cứu và những đơn vị tổng thể, về sự phân phối chúng trong một thời gian cũng như về sự thay đổi từ thời kỳ này tới thời kỳ khác được khẳng định qua tiêu thức thời gian. Ví dụ tổng số máy điện thoại có đến 31/12/2001 là 4.301.120, trong đó 70,01% máy cố định 29,09% máy di động.

3. Tiêu thức không gian:

Nêu phạm vi lãnh thổ bao trùm của đối tượng nghiên cứu và sự xuất hiện theo địa điểm của các đơn vị tổng thể. Những biểu hiện của nó được chỉ ra nhờ sự phân định về mặt quản lý hành chính hoặc theo điều kiện tự nhiên, phân vùng kinh tế ... Nghiên cứu thống kê theo tiêu thức không gian có ý nghĩa quan trọng trước hết là gắn với tiêu thức thực thể để quan sát sự phân phối về mặt lãnh thổ của các đơn vị tổng thể.

Các tiêu thức góp phần vào việc khẳng định đơn vị tổng thể cũng như tổng thể. Vì chúng nêu rõ các mặt và tính chất nhất định của đơn vị tổng thể cũng như của tổng thể. Nhờ đó có thể phân biệt đơn vị này với đơn vị khác cũng như tổng thể này với tổng thể kia.

1.2.5. Chỉ tiêu thống kê (gọi tắt là chỉ tiêu)

Nghiên cứu thống kê không chỉ phản ánh lượng và chất của hiện tượng kinh tế xã hội cá biệt mà còn phản ánh và chất của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể. Tính chất của các hiện tượng cá biệt được khái quát hóa trong chỉ tiêu thống kê. Do đó chỉ tiêu chỉ ra những mối quan hệ cần thiết, cái chung của tất cả các đơn vị hoặc của nhóm đơn vị. Ngoài ra chỉ tiêu còn phản ánh các mối quan hệ tồn tại khách quan, nhưng cũng không tự bộc lộ ra để hiểu trực tiếp là mối quan hệ. Phải điều tra mặt lượng của đơn vị cá biệt và từ đó phát hiện ý nghĩa theo số lượng của mối quan hệ bằng chỉ tiêu.

Chỉ tiêu thống kê có hai mặt: khái niệm và con số. Khái niệm của chỉ tiêu bao gồm định nghĩa và giới hạn về thực thể, thời gian và không gian của hiện tượng kinh tế xã hội. Mặt này chỉ rõ nội dung của chỉ tiêu thống kê. Con số của chỉ tiêu là trị số được phát hiện với đơn vị tính toán phù hợp. Nó nêu lên mức độ của chỉ tiêu. Theo nội dung, chỉ tiêu biểu hiện quy mô, cơ cấu, sự phát triển và mối quan hệ của hiện tượng số lớn trong điều kiện thời gian và địa điểm cụ thể.

Căn cứ vào nội dung có thể chia các chỉ tiêu thống kê thành hai loại: khối lượng và chất lượng. Chỉ tiêu khối lượng biểu hiện quy mô của tổng thể như số cán bộ công nhân viên, số máy điện thoại, khối lượng sản phẩm dịch vụ. Chỉ tiêu chất lượng biểu hiện trình độ phổ biến, mối quan hệ của tổng thể như giá thành sản phẩm dịch vụ. Việc phân loại này nhằm đáp ứng yêu cầu của một số phương pháp phân tích thống kê.

1.2.6 Thang đo trong thống kê

- **Thang đo định danh** (hay là đặt tên) là đánh số các biểu hiện cùng loại của một tiêu thức. Như giới tính biểu hiện “nam” được đánh số 1 và nữ đánh số 2. Giữa các con số ở đây không có quan hệ hơn, kém. Cho nên các phép tính với chúng đều vô nghĩa. Loại thang đo này dùng để đếm tần số của biểu hiện tiêu thức.

Ví dụ: thang đo định danh trong câu hỏi phỏng vấn:

Tình trạng hôn nhân của ông (bà):

1. Có gia đình
2. Độc thân
3. Ly dị
4. Trường hợp khác

- **Thang đo thứ bậc** cũng là thang đo định danh, nhưng giữa các biểu hiện tiêu thức có quan hệ thứ bậc, hơn, kém. Sự chênh lệch giữa các biểu hiện không nhất thiết phải bằng nhau, như huân chương có ba hạng: một, hai và ba. Hạng một hơn hạng hai, hạng hai hơn hạng ba. Trình độ văn hoá phổ thông có ba cấp: một, hai và ba. Cấp ba hơn cấp hai, cấp hai hơn cấp một. Con số có trị số lớn hơn không có nghĩa ở bậc cao hơn và ngược lại, mà do sự quy định. Thang đo loại này được sử dụng để tính toán đặc trưng chung của tổng thể một cách tương đối, trong một số trường hợp như tính cấp bậc bình quân của một doanh nghiệp, một đơn vị, bộ phận.

Ví dụ: trong câu hỏi phỏng vấn:

Thu nhập của ông (bà) hàng tháng:

1. < 4 triệu đồng
2. Từ 4 đến 7 triệu đồng
3. > 7 triệu đồng

- **Thang đo khoảng** là thang đo thứ bậc có các khoảng cách đều nhau. Có thể đánh giá sự khác biệt giữa các biểu hiện bằng thang đo loại này. Việc cộng trừ các con số có ý nghĩa, có thể tính các đặc trưng chung như số bình quân, phương sai. Yêu cầu có khoảng cách đều là đặt ra đối với thang đo, còn đối với biểu hiện của tiêu thức được đo không nhất thiết phải bằng nhau.

Ví dụ: trong câu hỏi phỏng vấn:

Đề nghị ông (bà) cho ý kiến về tầm quan trọng của mục tiêu đào tạo cho sinh viên đại học sau đây bằng cách khoanh tròn các số tương ứng trên thang đánh giá chỉ mức độ từ 1 đến 5 (1 = không quan trọng; 5 = rất quan trọng).

	Không quan	Bình thường	Rất quan trọng
--	------------	-------------	----------------

	trọng				
Đạo đức	1	2	3	4	5
Khả năng tư duy	1	2	3	4	5
Năng lực giải quyết vấn đề	1	2	3	4	5
Kiến thức chuyên môn	1	2	3	4	5
Sức khỏe	1	2	3	4	5
Làm việc nhóm	1	2	3	4	5
Làm việc độc lập	1	2	3	4	5

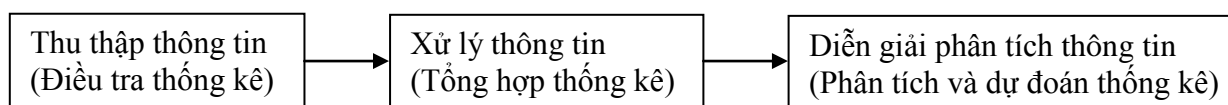
- **Thang đo tỷ lệ** là thang đo khoảng với một điểm không (0) tuyệt đối (điểm gốc) để có thể so sánh được tỷ lệ giữa các trị số đo. Với thang đo loại này, có thể đo lường các biểu hiện của tiêu thức như các đơn vị đo lường vật lý thông thường (kg, m,...) và thực hiện được tất cả các phép tính với trị số đo.

1.2.7. Hoạt động thống kê và quá trình nghiên cứu thống kê

Mục đích cuối cùng của một cuộc nghiên cứu thống kê là thu thập những thông tin định lượng về hiện tượng nghiên cứu trong điều kiện lịch sử cụ thể, trên cơ sở đó phát hiện, bản chất quy phát triển của hiện tượng, giải quyết được một vấn đề lý thuyết hoặc một yêu cầu nhất định của một thực tiễn.

Hoạt động thống kê là điều tra, báo cáo, tổng hợp, phân tích và công bố các thông tin phản ánh bản chất và tính quy luật của các hiện tượng kinh tế xã hội trong điều kiện không gian và thời gian cụ thể.

Các hoạt động thống kê đều phải trải qua một quá trình gồm nhiều giai đoạn, nhiều bước công việc kế tiếp nhau, có liên quan chặt chẽ với nhau. Có thể khái quát quá trình này bằng một sơ đồ sau:



Điều tra thống kê

Điều tra thống kê là việc tổ chức một cách khoa học với một kế hoạch thống nhất việc thu thập, ghi chép nguồn tài liệu ban đầu về hiện tượng nghiên cứu trong điều kiện cụ thể về thời gian và không gian.

Tổng hợp thống kê

Tổng hợp thống kê được hiểu là quá trình tập trung, chỉnh lý, hệ thống hóa các tài liệu thu được trong điều tra thống kê để làm cho các đặc trưng riêng biệt của từng đơn vị điều tra bước đầu chuyển thành những thông tin chung của toàn bộ hiện tượng nghiên cứu, làm cơ sở cho việc phân tích tiếp theo,...

Phân tích và dự đoán thống kê

Phân tích và dự đoán thống kê được hiểu là việc nêu lên một cách tổng hợp bản chất và tính quy luật của các hiện tượng và quá trình kinh tế xã hội số lớn trong điều kiện lịch sử cụ thể qua các biểu hiện về lượng và tính toán các mức độ trong tương lai nhằm đưa ra những căn cứ cho quyết định quản lý.

1.3 Các phương pháp trình bày dữ liệu thống kê

1.3.1 Bảng thống kê

Sau khi tổng hợp các tài liệu điều tra thống kê, muốn phát huy tác dụng của nó đối với giai đoạn phân tích thống kê, cần thiết phải trình bày kết quả tổng hợp theo một hình thức thuận lợi nhất cho việc sử dụng sau này. Có thể trình bày các kết quả tổng hợp bằng các hình thức: Bảng thống kê, đồ thị thống kê, bài viết,...

1. Ý nghĩa tác dụng của bảng thống kê

Bảng thống kê là một hình thức trình bày các tài liệu thống kê một cách có hệ thống, hợp lý và rõ ràng, nhằm nêu lên các đặc trưng về mặt lượng của hiện tượng nghiên cứu. Đặc điểm chung của tất cả các bảng thống kê là bao giờ cũng có những con số của từng bộ phận và chung có liên hệ mật thiết với nhau.

Bảng thống kê có nhiều tác dụng quan trọng trong mọi công tác nghiên cứu kinh tế nói chung và phân tích thống kê nói riêng. Các tài liệu trong bảng thống kê đã được sắp xếp lại một cách khoa học, nên có thể giúp ta tiến hành mọi việc so sánh đối chiếu, phân tích theo các phương pháp khác nhau, nhằm nêu lên sâu sắc bản chất của hiện tượng nghiên cứu. Nếu biết trình bày và sử dụng thích đáng các bảng thống kê, thì việc chứng minh vấn đề sẽ trở nên rất sinh động, có sức thuyết phục hơn cả những bài văn dài.

2. Cấu thành bảng thống kê

- *Về hình thức:* Bảng thống kê bao gồm các hàng ngang, cột dọc, các tiêu đề, tiêu mục và các tài liệu con số.

Các hàng ngang, cột dọc phản ánh quy mô của bảng thống kê. Các hàng ngang cột dọc cắt nhau tạo thành các ô dùng để điền các số liệu.

Tiêu đề của bảng thống kê phản ánh nội dung, ý nghĩa của bảng và của từng chi tiết trong bảng. Các tài liệu số thu thập được do kết quả tổng hợp thống kê được ghi vào các ô của bảng thống kê, mỗi con số phản ánh một đặc trưng về mặt lượng của hiện tượng nghiên cứu.

- *Về nội dung*: Bảng thống kê gồm hai phần: phần chủ đề và phần giải thích.

Phần chủ đề nói lên hiện tượng được trình bày trong bảng thống kê, tổng thể này được phân thành những đơn vị nào, bộ phận nào? Nó giải đáp những vấn đề: đối tượng nghiên cứu của bảng thống kê là những đơn vị nào, những loại hình gì?...

Phần giải thích gồm các chỉ tiêu giải thích các đặc điểm của đối tượng nghiên cứu, tức là giải thích phần chủ đề của bảng.

Cấu thành của bảng thống kê có thể biểu hiện bằng sơ đồ sau:

Phần chủ đề	Các chỉ tiêu giải thích			
(a)	(1)	(2)	(3)	(4)
Tên chủ đề (tên hàng)				

3. Các loại bảng thống kê

Căn cứ theo kết cấu của phần chủ đề, có thể chia làm ba loại bảng thống kê: Bảng giản đơn, bảng phân tổ và bảng kết hợp.

Bảng giản đơn:

Bảng giản đơn là loại bảng thống kê, trong đó phần chủ đề không phân tổ. Trong phần chủ đề của bảng giản đơn có liệt kê các đơn vị tổng thể, tên gọi các địa phương hoặc các thời gian khác nhau của quá trình nghiên cứu.

Bảng phân tổ:

Bảng phân tổ là loại bảng thống kê, trong đó đối tượng nghiên cứu ghi trong phần chủ đề được phân chia thành tổ theo một tiêu thức nào đó. Các bảng phân tổ là kết quả của việc áp dụng phương pháp phân tổ thống kê. Bảng phân tổ cho ta thấy rõ các loại hình kinh tế - xã hội tồn tại trong bản thân hiện tượng nghiên cứu, nêu lên kết cấu và biến động kết cấu của hiện tượng; trong nhiều trường hợp còn giúp ta phân tích được mối liên hệ giữa các hiện tượng.

Bảng kết hợp:

Bảng kết hợp là loại bảng thống kê, trong đó đối tượng nghiên cứu ghi trong phần chủ đề được phân tổ theo hai, ba,... tiêu thức kết hợp với nhau. Loại bảng kết hợp như trên giúp ta nghiên cứu được sâu sắc bản chất của hiện tượng, đi sâu vào kết cấu nội bộ của hiện tượng trong quá trình phát triển.

4. Những yêu cầu đối với việc xây dựng bảng thống kê

- Quy mô bảng thống kê không nên quá lớn (quá nhiều tổ hoặc quá nhiều chỉ tiêu giải thích). Khi có nhiều tiêu thức cần phân tổ có nhiều chỉ tiêu giải thích thì nên tách ra xây dựng

một số bảng thống kê.

- Các tiêu đề, tiêu mục trong bảng thống kê phải được ghi chính xác, đầy đủ, ngắn gọn và dễ hiểu.

- Các chỉ tiêu giải thích trong bảng thống kê cần được sắp xếp theo thứ tự hợp lý, phù hợp với mục đích nghiên cứu. Các chỉ tiêu có liên hệ với nhau nên sắp xếp gần nhau.

- Các ô trong bảng thống kê dùng để ghi các con số thống kê. Nếu không có số liệu để ghi vào một hoặc một số ô nào đó thì dùng các ký hiệu theo quy ước.

Trong bảng thống kê phải dùng đơn vị tính cụ thể cho từng chỉ tiêu. Nếu tất cả các số trong bảng có cùng đơn vị thì đơn vị tính ghi ở đầu bảng. Nếu các chỉ tiêu có đơn vị tính khác nhau thì đơn vị tính ghi ngay dưới tiêu mục.

Dưới bảng thống kê cần ghi rõ nguồn tài liệu sử dụng và các chi tiết cần thiết mà không thể hiện được trong bảng thống kê.

1.3.2. Đồ thị thống kê

1. Ý nghĩa và tác dụng của đồ thị

Đồ thị thống kê là các hình vẽ hoặc đường nét hình học dùng để miêu tả có tính chất quy ước các tài liệu thống kê.

Đồ thị thống kê có mấy đặc điểm sau;

- Đồ thị thống kê sử dụng con số kết hợp với hình vẽ, đường nét và màu sắc để trình bày và phân tích các đặc trưng số lượng của hiện tượng. Vì vậy người xem không mất nhiều công đọc con số mà vẫn nhận thức được vấn đề chủ yếu một cách dễ dàng, nhanh chóng.

- Đồ thị thống kê chỉ trình bày một cách khái quát các đặc điểm chủ yếu về bản chất và xu hướng phát triển của các hiện tượng.

Do các đặc điểm nêu trên, đồ thị thống kê có tính quần chúng, có sức hấp dẫn và sinh động, làm cho người hiểu biết ít về thống kê vẫn lĩnh hội được vấn đề chủ yếu một cách dễ dàng, đồng thời giữ được ấn tượng khá sâu đối với hiện tượng.

Phương pháp đồ thị thống kê được ứng dụng rộng rãi trong nghiên cứu, nhằm mục đích hình tượng hóa:

- Sự phát triển của hiện tượng qua thời gian
- Kết cấu và biến động của kết cấu qua hiện tượng
- Trình độ phổ biến của hiện tượng
- Sự so sánh giữa các mức độ của hiện tượng
- Mối liên hệ giữa các hiện tượng
- Tình hình thực hiện kế hoạch

Ngoài ra, đồ thị thống kê còn được coi là một phương tiện tuyên truyền, một công cụ dùng để biểu dương các kết quả sản xuất.

2. Các loại đồ thị thống kê

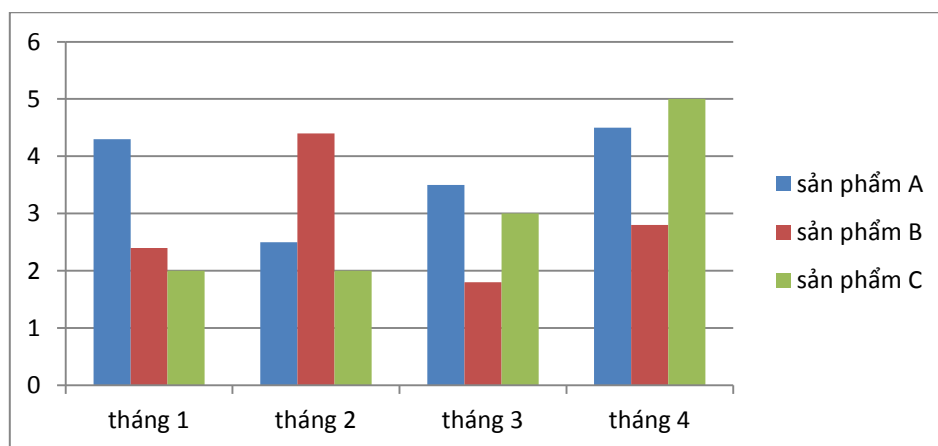
Căn cứ theo nội dung phản ánh, có thể phân chia đồ thị thống kê thành các loại sau:

- Đồ thị so sánh
- Đồ thị phát triển
- Đồ thị kết cấu
- Đồ thị hoàn thành kế hoạch
- Đồ thị liên hệ
- Đồ thị phân phối.

Căn cứ vào hình thức biểu hiện, có thể phân chia thành:

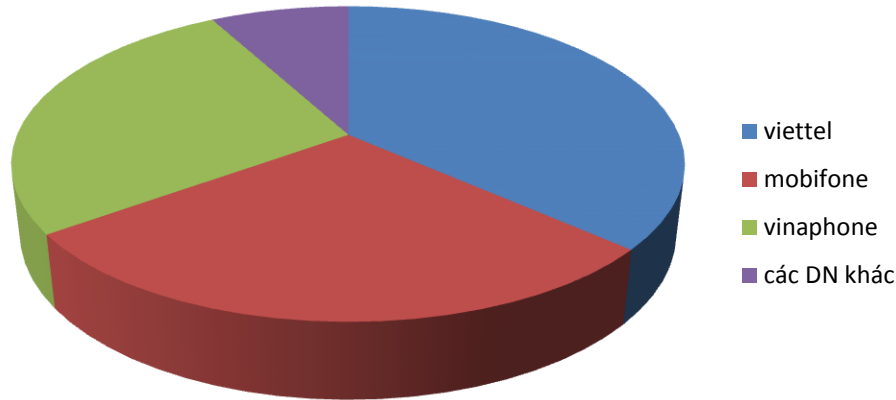
- Biểu đồ hình cột
- Biểu đồ tượng hình
- Biểu đồ diện tích
- Đồ thị đường gấp khúc

Ví dụ về biểu đồ hình cột:



Ví dụ về biểu đồ diện tích

Thị phần các doanh nghiệp thông tin di động trên thị trường thời điểm 6/2012



3. Yêu cầu đối với việc xây dựng đồ thị thống kê

Một đồ thị thống kê phải đảm bảo các yêu cầu: chính xác, dễ xem, dễ hiểu. Ngoài ra còn phải thể hiện tính thẩm mỹ của đồ thị. Cho nên khi xây dựng đồ thị thống kê cần thực hiện các yêu cầu sau:

- Lựa chọn loại đồ thị cho phù hợp với nội dung, tính chất các số liệu cần diễn đạt. Để đảm bảo những yêu cầu trên, cần chú ý đến các yếu tố của đồ thị, quy mô, các ký hiệu hình học hoặc các hình vẽ, hệ tọa độ, thang và tỷ lệ xích, phân giải thích.

- Xác định quy mô đồ thị cho thích hợp. Quy mô của đồ thị được quyết định bởi chiều dài, chiều cao và quan hệ tỷ lệ giữa 2 chiều đó. Quy mô của đồ thị to hay nhỏ còn phải căn cứ vào mục đích sử dụng. Quan hệ tỷ lệ giữa chiều cao và chiều dài của đồ thị, thông thường được dùng từ 1 : 1,33 đến 1 : 1,50.

- Các ký hiệu hình học hoặc hình vẽ quyết định dáng của đồ thị. Các ký hiệu hình học có nhiều loại như: các chấm, các đường thẳng hoặc cong, các hình cột, hình vuông, hình chữ nhật, hình tròn v.v... Các hình vẽ khác trên đồ thị cũng có thể thay đổi nhiều loại tùy tính chất của hiện tượng nghiên cứu. Việc lựa chọn các ký hiệu hình học hoặc hình vẽ của đồ thị là vấn đề quan trọng, vì mỗi hình có khả năng diễn tả riêng. Như khi cần biểu hiện kết cấu thành phần thời gian quay vòng toa xe có thể vẽ các hình cột hoặc các hình tròn (có chia thành các hình quạt).

- Hệ tọa độ giúp cho việc xác định chính xác vị trí các ký hiệu hình học trên đồ thị. Các đồ thị thống kê thường dùng hệ tọa độ vuông góc. Trên hệ tọa độ vuông góc, trục hoành thường được dùng để biểu thị thời gian, trục tung biểu thị trị số của chỉ tiêu. Trong trường hợp phân tích mối liên hệ giữa hai tiêu thức, thì tiêu thức nguyên nhân được để ở trục hoành, tiêu thức kết quả được ghi trên trục tung.

- Thang và tỷ lệ xích giúp cho việc tính chuyển các đại lượng lên đồ thị theo các khoảng

cách thích hợp. Người ta thường dùng các thang đường thẳng, được phân theo các trục tọa độ. Cũng có khi dùng thang đường cong như thang tròn (ở đồ thị hình tròn) được chia thành 360^0 . Các thang tỷ lệ có thể có khoảng cách bằng nhau hoặc không bằng nhau. Các thang tỷ lệ có các khoảng cách không bằng nhau chỉ dùng để biểu hiện các tốc độ khi khoảng biến thiên của các mức độ quá lớn mà người ta chỉ chú ý đến biến động tương đối của chúng.

- Phải giải thích tên đồ thị, các con số và ghi chú dọc theo thang tỷ lệ, các con số bên cạnh từng bộ phận của đồ thị, giải thích các ký hiệu quy ước ... cần được ghi rõ, gọn dễ hiểu.

TÀI LIỆU THAM KHẢO CHƯƠNG 1

1. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – Giáo trình **Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
2. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. TS. Hồ Sỹ Chi - **Thống kê doanh nghiệp**. NXB Tài chính, 2000
5. GS.TS. Phạm Ngọc Kiểm - **Giáo trình Thống kê doanh nghiệp**. NXB Lao động - Xã hội, 2012.
6. GS.TS. Phạm Ngọc Kiểm TS. Nguyễn Công Nhự - **Giáo trình Thống kê doanh nghiệp**. Nhà xuất bản Thống kê, 2007

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 1

1. Phân tích đối tượng nghiên cứu của thống kê học.
2. Trình bày các khái niệm thường dùng trong thống kê, ý nghĩa của các khái niệm này? Cho ví dụ?
3. Trình bày các loại thang đo trong thống kê. Cho ví dụ về việc sử dụng các loại thang đo này trong thực tế?
4. Trình bày những yêu cầu đối với việc xây dựng bảng thống kê.
5. Trình bày những yêu cầu đối với việc xây dựng đồ thị thống kê.

CHƯƠNG 2

CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG THỐNG KÊ

Mọi hiện tượng kinh tế - xã hội đều tồn tại trong điều kiện không gian, thời gian nhất định và mặt lượng của hiện tượng được biểu hiện ở các mức độ khác nhau. Ta có thể sử dụng các chỉ tiêu đo lường tính toán các mức độ này như số tuyệt đối, số tương đối, các đặc trưng đo lường khuynh hướng tập trung, các đặc trưng đo lường độ phân tán nhằm nêu lên đặc trưng phân phối của dãy số.

2.1. Số tuyệt đối

2.1.1. Khái niệm và ý nghĩa của số tuyệt đối

Trong thống kê số tuyệt đối biểu hiện quy mô, khối lượng hiện tượng trong điều kiện thời gian và không gian cụ thể. Số tuyệt đối trong thống kê có thể biểu hiện số đơn vị của tổng thể hay bộ phận (số doanh nghiệp, số công nhân) hoặc các trị số của 1 tiêu thức hay chỉ tiêu thống kê nào đó. (khối lượng sản phẩm dịch vụ, tổng chi phí sản xuất)...

Ví dụ năm 2016, số lao động của doanh nghiệp X là 400 người và doanh thu của doanh nghiệp này là 60 tỉ đồng, các con số trên đều là số tuyệt đối

Số tuyệt đối có ý nghĩa quan trọng đối với công tác quản lý kinh tế.

- Thông qua số tuyệt đối có thể nhận thức được cụ thể về quy mô, khối lượng của hiện tượng nghiên cứu, khả năng tiềm tàng và kết quả đạt được về sự phát triển kinh tế xã hội.
- Số tuyệt đối là cơ sở đầu tiên để phân tích thống kê và là cơ sở để tính các loại chỉ tiêu thống kê khác như số tương đối, số bình quân.
- Là căn cứ để xây dựng kế hoạch và chỉ đạo thực hiện kế hoạch.

2.1.2. Đặc điểm của số tuyệt đối

Số tuyệt đối trong thống kê bao giờ cũng phải mang một nội dung kinh tế nào đó trong thời gian và không gian cụ thể, có nghĩa là số tuyệt đối trong thống kê phải phản ánh được nội dung gì? của đơn vị nào? khi nào?

Số tuyệt đối trong thống kê phải được xác định qua điều tra và tổng hợp thống kê, nó không phải là con số lựa chọn tùy ý. Nó được xác định thông qua một hoặc một số phép tính toán.

2.1.3. Phân loại số tuyệt đối

1. Số tuyệt đối thời kỳ: Số tuyệt đối thời kỳ là số tuyệt đối phản ánh quy mô, khối lượng của hiện tượng trong một độ dài thời gian nhất định. Nó được hình thành thông qua sự tích lũy về lượng của hiện tượng trong suốt thời gian nghiên cứu. Các số tuyệt đối thời kỳ của cùng một chỉ tiêu có thể cộng lại với nhau để có trị số của thời kỳ dài hơn. Thời kỳ càng dài, trị số của chỉ tiêu càng lớn.

2. Số tuyệt đối thời điểm: Số tuyệt đối thời điểm là số tuyệt đối phản ánh mặt lượng của hiện tượng nghiên cứu tại một thời điểm nhất định. Với số tuyệt đối thời điểm thì không thể cộng lại với nhau. Vì khi cộng lại chúng không phản ánh được nội dung kinh tế nào. Số tuyệt đối thời điểm chỉ phản ánh tình hình của hiện tượng tại một thời điểm nào đó, trước và sau thời điểm trạng thái của hiện tượng có thể khác. Muốn có số tuyệt đối thời điểm chính xác phải quy định thời điểm hợp lý và phải tổ chức điều tra kịp thời.

2.1.4 Đơn vị tính số tuyệt đối

Trong thống kê, các số tuyệt đối đều có đơn vị tính cụ thể để biểu thị nội dung của hiện tượng nghiên cứu. Tùy theo tính chất của hiện tượng và mục đích nghiên cứu có thể sử dụng các đơn vị tính khác nhau.

- *Đơn vị tự nhiên:* Là đơn vị tính phù hợp với đặc điểm vật lý của hiện tượng, cái, con, chiếc... hoặc thời gian (phút, giờ, ngày...). Trong nhiều trường hợp phải dùng đơn vị kép như mật độ điện thoại (máy / 100 dân).

- *Đơn vị thời gian:* Thường dùng để tính lượng lao động hao phí sản xuất ra những sản phẩm dịch vụ không thể tổng hợp hoặc so sánh với nhau bằng đơn vị tính toán khác, hoặc những sản phẩm dịch vụ phức tạp do nhiều người thực hiện qua nhiều giai đoạn khác nhau.

- *Đơn vị tiền tệ (VNĐ, USD, EURO,...):* được sử dụng rộng rãi để biểu thị giá trị của sản phẩm dịch vụ. Nó giúp cho việc tổng hợp và so sánh nhiều loại sản phẩm có giá trị sử dụng và đo lường khác nhau. Tuy nhiên, do giá cả luôn thay đổi nên đơn vị tiền tệ không có tính chất so sánh được qua thời gian. Để khắc phục nhược điểm này phải dùng “giá cố định” ở thời gian nào đó.

2.2 Số tương đối

2.2.1. Khái niệm và ý nghĩa số tương đối

1. Khái niệm: Số tương đối trong thống kê là một loại chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng nghiên cứu cùng loại nhưng khác nhau về thời gian hoặc không gian, hoặc biểu diễn quan hệ tỷ lệ giữa hai chỉ tiêu thống kê khác loại nhưng có liên quan đến nhau.

2. Ý nghĩa: Trong thống kê số tương đối có ý nghĩa quan trọng. Nó là một trong những chỉ tiêu phân tích thống kê. Nếu như số tuyệt đối chỉ khái quát được về quy mô khối lượng của hiện tượng nghiên cứu, thì số tương đối cho phép phân tích đặc điểm của hiện tượng, nghiên cứu các hiện tượng trong mối quan hệ so sánh với nhau.

Trong thống kê số tương đối phản ánh các mặt sau:

- Số tương đối cho biết kết cấu của hiện tượng;
- Biểu hiện mối quan hệ giữa hai chỉ tiêu;
- Biểu hiện trình độ phát triển của hiện tượng (so sánh thực hiện giữa các năm);

- Trình độ phổ biến của hiện tượng (mật độ dân số, mật độ điện thoại,...);
- Giúp ta kiểm tra đánh giá tình hình thực hiện kế hoạch;
- Đi sâu phân tích so sánh đặc điểm của hiện tượng và giữ bí mật số tuyệt đối

2.2.2. Đặc điểm của số tương đối

Trong thống kê số tương đối không phải là con số trực tiếp thu được qua điều tra mà là kết quả so sánh giữa hai số tuyệt đối đã có. Vì vậy mỗi số tương đối phải có gốc để so sánh (gốc so sánh gọi là chỉ tiêu gốc). Tùy theo mục đích nghiên cứu mà chọn gốc so sánh cho phù hợp.

Trong thống kê số tương đối có thể biểu thị bằng lần hoặc phần trăm (%), phần nghìn (o/oo)

2.2.3. Các loại số tương đối

1. Số tương đối động thái (tốc độ phát triển, chỉ số phát triển)

Số tương đối động thái biểu hiện sự biến động về mức độ của hiện tượng nghiên cứu qua thời gian. Nó được xác định bằng cách so sánh 2 mức độ của chỉ tiêu cùng loại nhưng khác nhau về thời gian, được biểu hiện bằng số lần hoặc số phần trăm.

$$t_{dt} = \frac{y_1}{y_0} \text{ (lần)}, \text{ hoặc } t_{dt} = \frac{y_1}{y_0} 100 \text{ (\%)}$$

Trong đó: y_1 - mức độ của hiện tượng kỳ nghiên cứu (kỳ báo cáo);

y_0 - mức độ của hiện tượng kỳ gốc;

t_{dt} - Số tương đối động thái.

Ví dụ: vốn đầu tư của một doanh nghiệp năm 2015 là 250 tỉ đồng và năm 2017 là 300 tỉ đồng. Nếu đem so sánh vốn đầu tư năm 2017 và năm 2015 ta có số tương đối động thái:

$$\frac{300}{250} = 1,2 \text{ lần (hay 120\%)}$$

Như vậy, vốn đầu tư năm 2017 so với 2015 bằng 1,2 lần hay 120%. Trong thực tế số tương đối động thái này được gọi là tốc độ phát triển hay chỉ số phát triển.

Muốn có số tương đối động thái chính xác cần đảm bảo tính chất so sánh được giữa các mức độ của kỳ báo cáo và kỳ gốc. Tức là phải đảm bảo giống nhau về nội dung kinh tế, về phương pháp tính và đơn vị tính, về phạm vi và độ dài thời gian mà mức độ phản ánh.

2. Số tương đối kế hoạch

Số tương đối kế hoạch được dùng để lập và kiểm tra tình hình thực hiện kế hoạch về một chỉ tiêu nào đó. Có hai loại số tương đối kế hoạch:

+ Số tương đối hoàn thành kế hoạch biểu hiện quan hệ tỷ lệ giữa mức độ thực tế đạt được trong kỳ so với mức độ kế hoạch đã đề ra của chỉ tiêu kinh tế nào đó.

$$t_{htkh} = \frac{y_1}{y_{kh}} \quad \text{hoặc} \quad t_{htkh} = \frac{y_1}{y_{kh}} \times 100$$

Trong đó : y_{kh} - mức độ kế hoạch;

y_1 - mức độ thực tế của kỳ báo cáo;

t_{htkh} - Số tương đối kế hoạch.

+ Số tương đối nhiệm vụ kế hoạch biểu hiện mối quan hệ tỷ lệ giữa mức độ cần đạt tới của chỉ tiêu kinh tế nào đó trong kỳ kế hoạch so với mức độ thực tế của chỉ tiêu đó ở kỳ gốc. Như vậy số tương đối nhiệm vụ kế hoạch phản ánh mục tiêu cần đạt tới của đơn vị.

$$t_{nvkh} = \frac{y_{kh}}{y_o} \quad \text{hoặc} \quad t_{nvkh} = \frac{y_{kh}}{y_o} \times 100$$

Trong đó: t_{nvkh} - số tương đối nhiệm vụ kế hoạch;

Giữa số tương đối động thái, số tương đối hoàn thành kế hoạch, số tương đối nhiệm vụ kế hoạch của cùng một chỉ tiêu, cùng một thời gian có quan hệ như sau:

$$t_{dt} = t_{nvkh} \cdot t_{htkh}$$

Quan hệ này được vận dụng để tính mức độ chưa biết khi đã biết các mức độ kia.

3. Số tương đối kết cấu

Số tương đối kết cấu phản ánh tỷ trọng của mỗi bộ phận cấu thành trong một tổng thể. Qua chỉ tiêu này có thể phân tích được đặc điểm cấu thành của hiện tượng.

Số tương đối kết cấu được tính bằng cách so sánh trị số tuyệt đối của chỉ tiêu của từng bộ phận so với trị số tuyệt đối của chỉ tiêu của cả tổng thể. Nó thường được biểu hiện bằng lần hoặc số phần trăm.

$$t_{kc} = \frac{y_{bp}}{y_{tt}} \quad \text{hoặc} \quad t_{kc} = \frac{y_{bp}}{y_{tt}} \times 100$$

Trong đó : y_{bp} - trị số tuyệt đối của chỉ tiêu của bộ phận;

y_{tt} - trị số tuyệt đối của chỉ tiêu của tổng thể;

t_{kc} - Số tương đối kết cấu.

Ví dụ: tỉ trọng thuê bao trả sau của vinaphone trong tổng số thuê bao của mạng là:

$$\frac{2,6}{26} \cdot 100 = 10\%$$

Muốn tính số tương đối kết cấu chính xác phải phân biệt được các bộ phận có tính chất khác nhau trong tổng thể nghiên cứu, có nghĩa là phải phân tổ chính xác, tức là phải dựa vào phương pháp phân tổ thống kê.

4. Số tương đối so sánh

Số tương đối so sánh biểu hiện quan hệ so sánh giữa các hiện tượng cùng loại khác nhau về không gian. Số tương đối so sánh còn biểu hiện sự so sánh giữa các bộ phận trong cùng một tổng thể, khi so sánh người ta lấy một bộ phận nào đó làm gốc rồi đem các bộ phận khác so sánh với nó.

5. Số tương đối cường độ

Biểu hiện trình độ phổ biến của hiện tượng nghiên cứu trong điều kiện lịch sử nhất định, nó được xác định bằng cách so sánh chỉ tiêu của 2 hiện tượng khác nhau nhưng có liên quan với nhau. Mức độ của hiện tượng mà ta cần nghiên cứu trình độ phổ biến được đặt ở tử số, còn mức độ của hiện tượng có liên quan được đặt ở mẫu số. Hình thức biểu hiện của số tương đối cường độ là đơn vị kép do đơn vị tính của tử số và mẫu số hợp thành.

Ví dụ:

$$\text{Mật độ dân số} = \frac{\text{Tổng số dân(người)}}{\text{Diện tích đất đai(Km2)}} = \text{(đơn vị: người/Km2)}$$

2.3. Số bình quân

2.3.1. Khái niệm số bình quân

Số bình quân trong thống kê là loại chỉ tiêu biểu hiện mức độ điển hình theo một tiêu thức hoặc chỉ tiêu thống kê nào đó của một tổng thể bao gồm nhiều đơn vị cùng loại.

Một tổng thể thống kê thường bao gồm nhiều đơn vị, các đơn vị này cơ bản là giống nhau, nhưng biểu hiện cụ thể về mặt lượng theo các tiêu thức nào đó thường chênh lệch nhau. Những chênh lệch này do nhiều yếu tố, ngoài các nguyên nhân chung quyết định đặc điểm cơ bản của hiện tượng, còn có những nguyên nhân riêng ảnh hưởng tới từng đơn vị. Khi nghiên cứu tổng thể thống kê không chỉ nêu các đặc điểm riêng biệt của từng đơn vị, mà cần tìm một chỉ tiêu có khả năng mô tả một cách khái quát đặc điểm chung, điển hình nhất của hiện tượng. Mức độ đó chính là số bình quân.

2.3.2. Ý nghĩa của số bình quân

Số bình quân có tính tổng hợp và khái quát cao, chỉ cần một trị số là số bình quân nó có thể nêu lên mức độ chung nhất, phổ biến nhất của tiêu thức nghiên cứu. Nó không kể đến chênh lệch thực tế giữa các đơn vị tổng thể. Số bình quân không biểu hiện mức độ cá biệt mà biểu hiện mức độ tính chung nhất của các đơn vị tổng thể.

Do số bình quân biểu hiện đặc điểm chung của cả tổng thể nghiên cứu cho nên những nét riêng biệt có tính chất ngẫu nhiên của từng đơn vị cá biệt bị loại trừ, tức là số bình quân đã san bằng mọi chênh lệch giữa các đơn vị tổng thể về lượng biến của tiêu thức nghiên cứu. Sự san bằng này chỉ có ý nghĩa lớn khi tính cho một số khá nhiều đơn vị.

Số bình quân có vị trí quan trọng trong lý luận và trong công tác thực tế. Nó được sử dụng:

- *Nêu lên đặc điểm điển hình của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể;*

- *Dùng để so sánh các hiện tượng không cùng quy mô (có thể so sánh kết quả của hoạt động sản xuất kinh doanh của 2 đơn vị A,B, có thể 2 đơn vị này có số vốn khác nhau, có số lao động khác nhau nhưng có thể so sánh năng suất lao động bình quân của 2 đơn vị);*

- *Có thể dùng để nghiên cứu sự biến động của hiện tượng theo thời gian, nhằm thấy được xu hướng phát triển cơ bản của hiện tượng số lớn, nghĩa là của đại bộ phận các đơn vị.*

- *Số bình quân được dùng trong công tác lập kế hoạch, phân tích thống kê.*

2.3.3. Các loại số bình quân trong thống kê

Có rất nhiều loại số bình quân. Trong thống kê sử dụng loại nào cần căn cứ vào đặc điểm nguồn tài liệu sẵn có, mục đích nghiên cứu và ý nghĩa kinh tế xã hội của số bình quân. Số bình quân được phân thành hai nhóm:

- Nhóm 1: Các số bình quân được tính từ tất cả các lượng biến bằng cách chia đều tổng các lượng biến cho số đơn vị có lượng biến đó. Nhóm này bao gồm số trung bình cộng, trung bình nhân, trung bình điều hoà.

- Nhóm 2: Số bình quân chuyên dùng để mô tả đặc trưng phân phối các đơn vị của tổng thể theo trị số tiêu thức nghiên cứu. Nhóm này gồm có một và trung vị.

1. Số bình quân số học (Trung bình cộng)

Số bình quân số học còn gọi là số trung bình cộng được xác định bằng cách đem chia tổng lượng biến của tiêu thức cho tổng số đơn vị tổng thể.

* *Số trung bình cộng giản đơn*

- Nó được sử dụng trong trường hợp mỗi lượng biến chỉ xuất hiện một lần. Công thức tính:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Trong đó : x_i - là lượng biến của tiêu thức của đơn vị thứ i ;

\bar{x} - Số trung bình của tổng thể;

n - số đơn vị tổng thể.

Ví dụ: tính năng suất lao động bình quân của một tổ công nhân có 6 người trong đó những người này sản xuất được số sản phẩm tương ứng là: 50, 55, 60, 65, 70, 72.

Theo công thức trên:

$$\bar{x} = \frac{50 + 55 + 60 + 65 + 70 + 72}{6} = 62 \text{ (sản phẩm)}$$

* *Số trung bình cộng gia quyền (số trung bình số học có trọng số, số bình quân gia quyền)*

Khi mỗi lượng biến xi xuất hiện nhiều lần, tức là có tần số (f_i) khác nhau, muốn xác định số trung bình cộng cần nhân mỗi lượng biến xi với tần số f_i tương ứng, sau đó cộng lại và đem chia cho số đơn vị tổng thể:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_m f_m}{f_1 + f_2 + \dots + f_m} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i}$$

Trong đó: $\sum_{i=1}^m f_i$ - Số đơn vị tổng thể

x_i - Là lượng biến i;

f_i - tần số của tổ i;

m - Là số tổ;

$x_i f_i$ - Gọi là gia quyền.

Ví dụ: Tính năng suất lao động bình quân của công nhân theo tài liệu sau:

Năng suất lao động (x_i)	Số công nhân (f_i)	$x_i f_i$
50	3	150
55	5	275
60	10	600
65	12	780
70	7	490

72	3	216
Cộng	40	2511

$$\bar{x} = \frac{(50 * 3) + (55 * 5) + (60 * 10) + (65 * 12) + (70 * 7) + (72 * 3)}{3 + 5 + 10 + 12 + 7 + 3}$$

$$= 62,8 \text{ (sản phẩm)}$$

* Các dạng đặc biệt của số trung bình cộng gia quyền

- Tính số trung bình cộng gia quyền từ dãy số lượng biến có khoảng cách tở.

Trong trường hợp mỗi tở có một phạm vi lượng biến, cần có lượng biến đại diện để làm căn cứ tính toán, và lượng biến đại diện là các trị số giữa:

x'_i là trị số giữa của tở thứ i

$$x'_i = \frac{x_{i \min} + x_{i \max}}{2}$$

$x_{i \min}$, $x_{i \max}$ là giới hạn dưới, giới hạn trên của tở i

$$\bar{x} = \frac{\sum_{i=1}^m x'_i f_i}{\sum_{i=1}^m f_i}$$

- Tính số trung bình cộng chung của các số trung bình cộng tở.

Trong trường hợp này số trung bình cộng chung sẽ là số trung bình cộng gia quyền của các số trung bình cộng tở trong đó quyền số là số đơn vị của mỗi tở:

$$\bar{x} = \frac{\sum_{i=1}^m \bar{x}_i f_i}{\sum_{i=1}^m f_i}$$

Trong đó: \bar{x}_i - là số trung bình cộng của tiêu thức của tở thứ i ;

f_i - là tần số của tở thứ i ;

m - là số tở.

- Tính số trung bình cộng gia quyền khi biết tỷ trọng và số đơn vị tổng thể của mỗi tở :

d_i - tỷ trọng của tở thứ i trong tổng thể : $d_i = \frac{f_i}{\sum f_i}$

$$\bar{x} = \frac{\sum_{i=1}^m x_i \cdot d_i}{1}$$

Trong đó: x_i - Là lượng biến của tiêu thức của tổ có tỷ trọng d_i

2. Số bình quân điều hoà

Được sử dụng trong trường hợp khi không có số liệu thống kê về số đơn vị tổng thể nhưng có số liệu thống kê về tổng các lượng biến của tiêu thức. Công thức tính:

$$\bar{x} = \frac{\sum_{i=1}^m M_i}{\sum_{i=1}^m \frac{M_i}{x_i}}$$

Trong đó: x_i - Là lượng biến thứ i ;

$M_i = x_i \cdot f_i$ - Là tổng lượng biến của tiêu thức của tổ thứ i ;

m - Là số tổ.

Số bình quân tính theo công thức trên gọi là số bình quân điều hoà gia quyền.

Các trường hợp đặc biệt:

- Nếu các tổng lượng biến của tiêu thức có giá trị bằng nhau, nghĩa là $M_1=M_2=M_3...= M$ thì:

$$\bar{x} = \frac{mM}{M \sum_{i=1}^m \frac{1}{x_i}} = \frac{m}{\sum_{i=1}^m \frac{1}{x_i}}$$

- Nếu các tổng lượng biến tiêu thức được xác định dưới tỷ lệ kết cấu, nghĩa là không có tài liệu về M_i về số tuyệt đối, mà chỉ biết các tỷ lệ $M_i / \sum M_i$ thì số bình quân điều hoà được xác định như sau:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}} = \frac{1}{\frac{M_1}{\sum M_i} * \frac{1}{x_1} + \frac{M_2}{\sum M_i} * \frac{1}{x_2} + \dots + \frac{M_m}{\sum M_i} * \frac{1}{x_m}} = \frac{1}{\sum_{i=1}^m \frac{d_i}{x_i}}$$

Trong đó: $d_i' = \frac{M_i}{\sum M_i}$

- Nếu di tính bằng % với $\sum d_i = 100$ thì:

$$\bar{x} = \frac{100}{\sum \frac{d_i}{x_i}}$$

3. Số bình quân nhân

Số bình quân nhân được sử dụng trong trường hợp khi các lượng biến có quan hệ tích số với nhau, thường được sử dụng để tính tốc độ phát triển trung bình của chỉ tiêu thống kê nào đó. Có hai cách tính :

- *Số bình quân nhân giản đơn:*

Được sử dụng trong trường hợp mỗi lượng biến xi chỉ xuất hiện một lần:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

Ví dụ : Tính tốc độ phát triển bình quân hàng năm về chỉ tiêu doanh thu của một đơn vị trong giai đoạn từ năm 2010-2016. Nếu biết:

+ Tốc độ phát triển về doanh thu:

Năm 2011 so với 2010 là 108 %

Năm 2012 so với năm 2011 là 109%

Năm 2013 so với năm 2012 là 106%

Năm 2014 so với năm 2013 là 108%

Năm 2015 so với năm 2014 là 110%

Năm 2016 so với năm 2015 là 109%

+ Tốc độ phát triển trung bình hàng năm về doanh thu được xác định theo công thức:

$$\bar{x} = \sqrt[6]{\prod_{i=1}^6 x_i} = \sqrt[6]{1,08 * 1,09 * 1,06 * 1,08 * 1,10 * 1,09} = 1,08326 \text{ (lần) hay } 108,326 \%$$

- *Số bình quân nhân gia quyền:*

Được dùng trong trường hợp mỗi lượng biến xi xuất hiện nhiều lần, tức là có tần số fi khác nhau:

$$\bar{x} = \sqrt[m]{\prod_{i=1}^m x_i^{f_i}}$$

Trong đó: xi - là lượng biến thứ i;

fi - là tần số của tổ thứ i;

m - là số tổ .

4. Mốt

a. *Khái niệm:* Một là biểu hiện của tiêu thức được gặp nhiều nhất trong tổng thể hay trong một dãy số phân phối. Như vậy, trong dãy số lượng biến Một là lượng biến có tần số lớn nhất. Trị số của Một không phụ thuộc vào lượng biến của tiêu thức mà phụ thuộc vào sự phân phối các đơn vị trong tổng thể.

Ví dụ: Có tài liệu về số lao động của 50 doanh nghiệp như sau:

Bảng 2.1

Số lao động (x_i)	Số doanh nghiệp (f_i)
15	4
18	3
21	7
24	9
27	11
28	8
30	7
32	1
Cộng	50

Theo định nghĩa trên có thể xác định Một là 27 lao động vì lượng biến này có tần số lớn nhất ($f_{i\max}=11$)

b. *Phương pháp xác định một:*

Phân biệt các trường hợp sau:

* *Đối với dãy số lượng biến không có khoảng cách tổ:* Trường hợp này một là lượng biến có tần số lớn nhất.

* *Đối với dãy số có khoảng cách tổ đều nhau:* Muốn tìm Một trước hết phải tìm tổ chứa một, tổ chứa một là tổ có tần số lớn nhất, sau đó tính trị số gần đúng của một theo công thức sau:

$$M_0 = x_{M_0\min} + h_{M_0} * \frac{f_{M_0} - f_{M_0-1}}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})}$$

Trong đó: $x_{M_0\min}$ - là giới hạn dưới của tổ chứa Một;

h_{M_0} - là khoảng cách của tổ có chứa Một;

$f_{M_0}, f_{M_0-1}, f_{M_0+1}$, là tần số của tổ chứa Một, tổ đứng trước tổ chứa Một, và tổ đứng sau tổ chứa Một.

Ví dụ có tài liệu khảo sát 100 doanh nghiệp về tình hình thực hiện doanh thu như sau:

Bảng 2.2

% thực hiện doanh thu	Số đơn vị
75-85	6
85-95	8
95-105	20
105-115	24
115-125	18
125-135	11
135-145	7
145-155	6
Cộng	100

Tổ có % thực hiện doanh thu từ 105-115% là tổ chứa một vì tổ này có tần số lớn nhất. Thay các số liệu vào công thức ta tính được Một.

$$M_0 = 105 + 10 * \frac{24 - 20}{(24 - 20) + (24 - 18)} = 109\%$$

* Trường hợp dãy số lượng biến có khoảng cách tổ không đều nhau, tổ chứa một là tổ có mật độ phân phối lớn nhất, tức là xung quanh lượng biến ấy chứa nhiều tần số nhất. Mật độ phân phối được xác định bằng cách đem chia tần số cho khoảng cách tổ tương ứng:

$$d_i = f_i / h_i$$

Trong đó d_i - Mật độ phân phối tổ thứ i ;

f_i - Tần số của tổ thứ i ;

h_i - Khoảng cách tổ i .

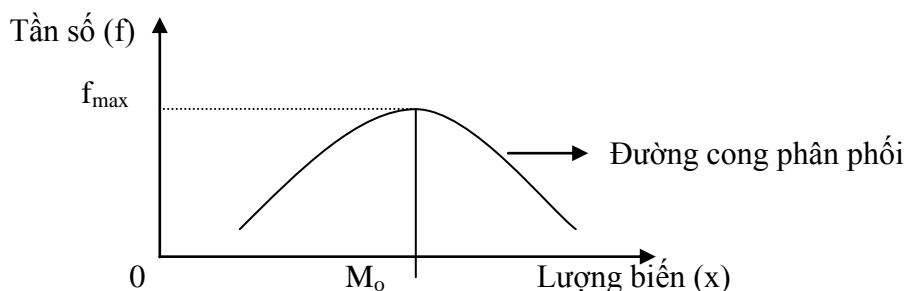
Khi đó Một được xác định theo công thức:

$$M_0 = x_{M_0 \min} + h_{M_0} * \frac{d_{M_0} - d_{M_0-1}}{(d_{M_0} - d_{M_0-1}) + (d_{M_0} - d_{M_0+1})}$$

c. Ý nghĩa của Mốt

Mốt biểu hiện mức độ phổ biến nhất của hiện tượng nghiên cứu, đồng thời bản thân nó lại không san bằng, bù trừ chênh lệch giữa các lượng biến, cho nên Mốt là chỉ tiêu có thể dùng bổ sung hoặc thay thế cho việc tính số trung bình khi cần thiết, trong trường hợp tính số trung bình gặp khó khăn.

Ngoài ra, mốt còn là một trong những mức độ để nêu lên đặc trưng phân phối của dãy số như nếu tính được số trung bình, mốt, trung vị có giá trị bằng nhau thì dãy số có phân phối chuẩn.



5. Số trung vị

a. Khái niệm

Số trung vị là lượng biến của tiêu thức của đơn vị đứng ở vị trí giữa trong dãy số lượng biến, chia dãy số lượng biến thành hai phần (phần trên và phần dưới số trung vị), mỗi phần có cùng có một số đơn vị tổng thể bằng nhau.

b. Phương pháp tính số trung vị:

+ Đối với dãy số lượng biến không có khoảng cách tổ.

Có hai trường hợp:

- Nếu số đơn vị của tổng thể lẻ: $n = 2k+1$, nghĩa là dãy số lượng biến:

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k+1}$$

Thì số trung vị sẽ là lượng biến của đơn vị đứng ở vị trí thứ $k+1$ tức là lượng biến x_{k+1} . Ký hiệu số trung vị là Me . Ta có

$$Me = x_{k+1}$$

- Nếu số đơn vị tổng thể chẵn ($n = 2k$) nghĩa là dãy số lượng biến:

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k}$$

Thì số trung vị được xác định căn cứ vào lượng biến của hai đơn vị đứng ở vị trí giữa nhất (đơn vị thứ k và k+1)

$$Me = \frac{x_k + x_{k+1}}{2}$$

+ Đối với dãy số lượng biến có khoảng cách tổ.

Muốn xác định số trung vị đầu tiên cần xác định tổ chứa trung vị, tức là tổ có chứa lượng biến của đơn vị ở vị trí giữa trong tổng số các đơn vị. Tổ có chứa số trung vị - đó là tổ đầu tiên có tần số tích lũy lớn hơn hoặc bằng nửa số đơn vị tổng thể, tức là tổ có tần số tích lũy bằng $(\sum f_i + 1)/2$. Sau đó tính giá trị gần đúng của số trung vị theo công thức sau:

$$Me = x_{M_e \min} + h_{M_e} * \frac{\sum \frac{f_i}{2} - S_{M_e - 1}}{f_{M_e}}$$

Trong đó: $x_{M_e \min}$: Là giới hạn dưới của tổ có số trung vị;

h_{M_e} - Khoảng cách tổ có chứa số trung vị;

$S_{M_e - 1}$ - Tần số tích lũy của tổ đứng trước tổ có số trung vị ;

f_{M_e} - Tần số của tổ chứa số trung vị .

Ví dụ: Có tài liệu về phân tổ công nhân theo NSLĐ của một đơn vị như sau:

Bảng 2.3

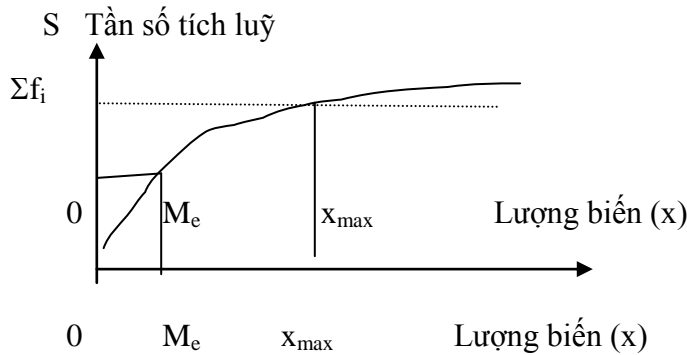
NSLĐ (Triệu đồng/người)	Số CN (người)	Tần số tích lũy (S)
20-22	5	5
22-24	9	14
24-26	20	34
26-28	10	44
28-30	6	50
Cộng	50	-

Ta có $(\sum f_i + 1)/2 = (50 + 1) / 2 = 25,5$, như vậy số trung vị rơi vào tổ 24-26. Do đó số trung vị năng suất lao động là :

$$M_e = 24 + 2 * \frac{\sum \frac{50}{2} - 14}{20} = 25,1 \text{ (Triệu đồng/ng)}.$$

c. Ý nghĩa

Cũng như một, số trung vị có khả năng nêu lên mức độ điển hình của tổng thể mà không san bằng, bù trừ chênh lệch giữa các lượng biến. Vì vậy có thể bổ sung hoặc thay thế số trung bình cộng khi không có đủ các lượng biến để tính. Việc tính số trung vị cũng thường được dùng để nêu lên đặc trưng của dãy số phân phối.



2.4. Chỉ tiêu đánh giá độ biến thiên của tiêu thức

2.4.1. Ý nghĩa nghiên cứu

Khi xét số trung bình, số trung vị, một người ta mới chỉ xét đến tính chất điển hình của toàn bộ tổng thể nghiên cứu mà chưa xét đến từng đơn vị cá biệt và sự chênh lệch thực tế giữa các lượng biến (có khi chênh lệch này là đáng kể hoặc có khi là không đáng kể). Do vậy ngoài việc tính mức độ trung bình, cần phải đánh giá độ biến thiên của tiêu thức nhằm:

- + Đánh giá tính chất đại biểu của số trung bình. Nếu độ biến thiên của tiêu thức càng nhiều thì tính chất đại biểu của số trung bình càng ít và ngược lại.
- + Trong phân tích hoàn thành kế hoạch các chỉ tiêu đánh giá độ biến thiên của tiêu thức giúp ta thấy rõ chất lượng công tác và nhịp độ hoàn thành kế hoạch chung và từng bộ phận trong doanh nghiệp.
- + Các chỉ tiêu đánh giá độ biến thiên của tiêu thức giúp ta nghiên cứu độ tương quan và tính sai số trong điều tra chọn mẫu.

2.4.2. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức

1. Khoảng biến thiên (R) hay còn gọi là toàn cự

Khoảng biến thiên là độ chênh lệch giữa lượng biến lớn nhất(xmax) và lượng biến nhỏ nhất (xmin) của tiêu thức nghiên cứu:

$$R = x_{\max} - x_{\min}$$

Khoảng biến thiên R càng nhỏ thì tính chất đồng đều của tổng thể nghiên cứu càng cao và ngược lại. Chỉ tiêu này giúp xác định nhanh chóng chênh lệch giữa đơn vị tiên tiến và đơn vị lạc hậu. Tuy nhiên nhược điểm của chỉ tiêu này là nó chỉ phụ thuộc vào lượng biến nhỏ nhất và lớn nhất của dãy số, không xét đến các lượng biến khác, cho nên nhiều khi dẫn đến những kết luận không hoàn toàn chính xác.

Ví dụ: Doanh thu tại 2 bưc cục như sau (triệu đồng)

Bưc cục A: 40; 50; 60; 70; 80

Bưc cục B : 58; 59; 60; 61; 62

Doanh thu bình quân của mỗi bưc cục là 60, nhưng thực ra 2 bưc cục này không đồng đều về doanh thu. Doanh thu của bưc cục A chênh lệch nhiều hơn so với bưc cục B. Để đánh giá mức độ biến thiên của doanh thu, qua đó đánh giá tính chất đại biểu của số bình quân, ta tính khoảng biến thiên của 2 bưc cục

$$R_A = 80 - 40 = 40 \text{ triệu đồng}$$

$$R_B = 62 - 58 = 4 \text{ triệu đồng}$$

Kết quả cho thấy R_A lớn hơn R_B , có nghĩa là độ biến thiên tiêu thức của bưc cục A lớn hơn và vì thế tính chất đại biểu của số bình quân bưc cục A thấp hơn.

2. Độ lệch tuyệt đối trung bình (\bar{d})

Độ lệch tuyệt đối trung bình là số trung bình cộng của các độ lệch tuyệt đối giữa các lượng biến và số trung bình cộng của các lượng biến đó.

Công thức tính:

- Khi lượng biến xuất hiện một lần:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

- Khi lượng biến xuất hiện nhiều lần:

$$\bar{d} = \frac{\sum_{i=1}^m |x_i - \bar{x}| f_i}{\sum_{i=1}^m f_i}$$

Trong đó: \bar{d} - Độ lệch tuyệt đối trung bình;

x_i - Các lượng biến của đơn vị thứ i (hoặc tổ i);

\bar{x} - Trung bình cộng của các lượng biến xi.

f_i - Tần số của tổ thứ i.

Khi độ lệch tuyệt đối trung bình càng nhỏ càng nhỏ thì tính chất đồng đều của tổng thể càng lớn, tính chất đại biểu của số trung bình số học càng cao và ngược lại. Độ lệch tuyệt đối trung bình có ưu điểm hơn khoảng biến thiên vì số đo này lưu ý đến tất cả các lượng biến trong dãy số.

3. Phương sai

Phương sai là số trung bình cộng của bình phương độ lệch giữa các lượng biến x_i với số trung bình cộng của các lượng biến đó (\bar{x}).

Công thức tính:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad \text{hoặc} \quad \sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum_{i=1}^m f_i}$$

Trong đó: σ^2 - là phương sai;

Phương sai càng nhỏ thì tính chất đồng đều của tổng thể càng cao, tính chất đại biểu của số trung bình số học càng cao và ngược lại.

Qua công thức tính phương sai cho thấy sự khác nhau về dấu giữa các độ lệch đã được khắc phục bằng cách tính bình phương. Song cũng chính vì thế mà trị số bị khuếch đại, đơn vị tính không phù hợp với thực tế. Để khắc phục nhược điểm này đưa ra chỉ tiêu độ lệch tiêu chuẩn.

4. Độ lệch tiêu chuẩn (σ)

Độ lệch tiêu chuẩn là căn bậc 2 của phương sai.

Công thức tính:
$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

hoặc
$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^m f_i}}$$

Chỉ tiêu độ lệch tiêu chuẩn là chỉ tiêu hoàn thiện nhất và thường dùng nhất trong nghiên cứu thống kê để đánh giá độ biến thiên của tiêu thức. Tuy nhiên việc tính toán là khá phức tạp.

5. Hệ số biến thiên (V)

Hệ số biến thiên là số tương đối so sánh giữa độ lệch tuyệt đối trung bình (\bar{d}) hoặc độ lệch tiêu chuẩn (σ) với số trung bình số học của của các lượng biến đó.

$$V = \frac{\bar{d}}{x} 100 \quad \text{và} \quad V = \frac{\sigma}{x} 100$$

Chỉ tiêu hệ số biến thiên cho phép so sánh 2 tổng thể không cùng quy mô, hoặc so sánh giữa 2 chỉ tiêu thống kê khác nhau.

2.4.3 Các phương pháp tính phương sai

1. Các tính chất toán học của phương sai

Tính chất 1: Phương sai của một hằng số bằng 0

$$\sigma^2_c = 0$$

Tính chất 2: Nếu tất cả các lượng biến x_i được cộng (hoặc trừ) đi một số x_0 bất kỳ thì phương sai tính ra không đổi

$$\sigma^2_{(x_i \pm x_0)} = \sigma^2_x$$

Trong đó: $\sigma^2_{(x_i \pm x_0)}$ - Phương sai của các lượng biến $x_i \pm x_0$

σ^2_x - Phương sai của các lượng biến x_i

Tính chất 3: Nếu tất cả các lượng biến x_i chia cho cùng một số d thì phương sai tính ra sẽ giảm đi d^2 lần.

$$\sigma^2_x = d^2 \cdot \sigma^2_{x/d}$$

Trong đó: $\sigma^2_{x/d}$ - Phương sai các lượng biến x_i đã chia cho cùng một số d .

Tính chất 4: Độ lệch bình phương bình quân giữa các lượng biến x_i với số bình quân cộng x bao giờ cũng nhỏ hơn độ lệch bình phương bình quân giữa các lượng biến x_i với một trị số A bất kỳ khác số bình quân cộng x . Nghĩa là phương sai có tính chất cực tiểu

$$\sigma^2_A = \sigma^2_x + (\bar{x} - A)^2$$

Trong đó: σ^2_A - Độ lệch bình phương bình quân giữa các lượng biến x_i với trị số A

2. Các phương pháp tính phương sai

Như trên đã nêu công thức tính phương sai theo định nghĩa. Song khi dãy số lượng biến có các giá trị lớn thì tính toán sẽ không thích hợp. Do vậy có thể tính phương sai bằng các phương pháp khác.

a. *Tính phương sai theo công thức biến đổi*: Cách tính này được suy ra từ công thức định nghĩa.

$$\sigma^2 = \overline{x^2} - (\bar{x})^2$$

$$\overline{x^2} = \left[\frac{\sum x_i f_i}{\sum f_i} \right]^2$$

Trong đó: $\overline{x^2}$ - Bình quân cộng của các bình phương x_i^2

$(\bar{x})^2$ - Bình phương của số bình quân cộng \bar{x}

b. *Tính phương sai bằng đổi biến*: Khi tài liệu phân tổ phức tạp, việc tính phương sai bằng công thức định nghĩa cũng như công thức biến đổi vẫn dài dòng và tốn kém thời gian. Để khắc phục có thể tính bằng phương pháp đổi biến. Phương pháp này dựa trên cơ sở một số tính chất toán học của phương sai.

Quá trình tính toán phương sai bằng phương pháp đổi biến tiến hành như sau:

+ Tạo biến mới

$$x' = \frac{x_i - x_0}{d}$$

Trong đó: x_0 và d là các hằng số tự chọn $\neq 0$

+ Tính phương sai theo biến mới (bằng công thức định nghĩa hay công thức biến đổi)

$$\sigma^2 = \frac{\sum (x'_i - \bar{x}')^2 f_i}{\sum f_i}$$

$$\sigma^2 = \overline{x'^2} - (\bar{x}')^2$$

+ Tính phương sai theo biến cũ:

$$\sigma^2 = \frac{1}{d^2} \sigma_x^2$$

do đó:

$$\sigma^2 = d^2 \cdot \sigma_{x'}^2$$

c. *Phương sai tiêu thức chất lượng*: Trong số các tiêu thức chất lượng thường gặp loại tiêu thức mà biểu hiện cụ thể của nó chỉ có 2 trường hợp: có hoặc không, nghĩa là ở một số đơn vị tổng thể có biểu hiện đặc tính quan tâm, còn một số đơn vị còn lại thì không biểu hiện.

Để xây dựng công thức tính số bình quân và phương sai cho loại tiêu thức này ta quy ước một số ký hiệu:

$x_1 = 1$ khi đơn vị quan sát có biểu hiện đặc tính quan tâm.

$x_2 = 0$ khi đơn vị quan sát không biểu hiện

p - Tỷ trọng của bộ phận có biểu hiện

q - Tỷ trọng của bộ phận không có biểu hiện.

p và q là những số tương đối nhỏ hơn 1, giữa chúng có mối quan hệ $p + q = 1$. Xác định số bình quân và phương sai căn cứ vào công thức định nghĩa.

+ Số bình quân cộng

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(1 \times p) + (0 \times q)}{p + q} = p$$

+ Phương sai

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \frac{(1 - p)^2 + (0 - q)^2}{p + q} = pq$$

2.5. Phân phối trong thống kê

2.5.1. Một số phân phối lý thuyết

1. Phân phối nhị thức

a. Công thức

Giả sử thực hiện m phép thử độc lập. Trong mỗi phép thử biến cố A xuất hiện với xác suất $P(A) = p$. Khi đó xác suất để biến cố A xuất hiện đúng x lần trong m phép thử được xác định bằng công thức:

$$P(x) = C_m^x p^x (1 - p)^{m-x}$$

Trong đó: x - Các đại lượng ngẫu nhiên có thể nhận các giá trị 0, 1, 2, 3, ... m .

$(1 - p)$ - Xác suất không xuất hiện biến cố A trong mỗi phép thử $1 - p = q$.

C_m^x - Các hệ số của phân phối nhị thức (tổ hợp chập x của tập hợp gồm m phần tử). Nó được xác định:

$$C_m^x = \frac{m!}{x! (m - x)!}$$

$$\text{Nhu vậy: } P(x) = \frac{m!}{x! (m - x)!} p^x q^{m-x}$$

b. Tính chất (không chứng minh)

- Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m P(x) = 1$$

- Nếu $p = q = 0,5$ có phân phối nhị thức đối xứng. Nếu $p \neq q$ có phân phối nhị thức không đối xứng.

- Các giá trị của hệ số nhị thức C_m^x luôn có tính chất đối xứng và có thể biểu diễn dưới dạng "tam giác" Pascal.

Các giá trị của C_m^x được tính sẵn và ghi trong bảng

c. Các tham số đặc trưng

- Số bình quân: $\bar{x} = mp$

- Phương sai: $\sigma_x^2 = m p q$

- Độ lệch chuẩn: $\sigma_x = \sqrt{m p q}$

2. Phân phối Poisson

Trong phân phối nhị thức khi m lớn, việc xác định các giá trị P(x) rất tốn kém thời gian. Để khắc phục tình trạng đó, có thể sử dụng phân phối Poisson thay thế. Điều kiện thay thế một cách hợp lý là xác suất p hoặc q khá bé và m tương đối lớn.

a. Công thức

Đặt $\lambda = mp$ và lúc đó xác suất để biến cố A nào đó xuất hiện đúng x lần có thể xác định bằng công thức:

$$P(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Trong đó: $e = 2,718 \dots$ Cơ số của logarit

$x = 0, 1, 2 \dots$ Đại lượng ngẫu nhiên rời rạc

b. Các tham số

- Số bình quân: $\bar{x} = \lambda = mp$

- Phương sai: $\sigma_x^2 = \lambda$

- Độ lệch chuẩn: $\sigma_x = \sqrt{\lambda}$

c. Tính chất

Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m P(x) = 1$$

Việc xác định các giá trị P (x) đã được ghi sẵn .

3. Phân phối chuẩn

Phân phối chuẩn sử dụng để xét các đại lượng ngẫu nhiên liên tục.

a. Phương trình đường cong phân phối chuẩn

Phân phối chuẩn là một phân phối liên tục. Trên đồ thị các tung độ của đường cong được xác định:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$

Trong đó: $\pi = 3,14159$; $e = 2,718$;

x - Đại lượng ngẫu nhiên liên tục có thể lấy mọi giá trị từ $-\infty$ đến $+\infty$

\bar{x} - Số bình quân ; σ_x - Độ lệch chuẩn.

Đường cong f(x) có dạng hình chuông đối xứng. Tại giá trị $x = \bar{x}$ tung độ đạt giá trị cực đại.

$$f(x) = f(\bar{x}) = \frac{1}{\sigma_x \sqrt{2\pi}} e^0 = \frac{1}{\sigma_x \sqrt{2\pi}}$$

Đỉnh của hình chuông là P ứng với tung độ cực đại. Ở hai bên của P tung độ giảm dần ; lúc đầu giảm ít, sau giảm nhiều, đến 2 điểm đối xứng ứng với hoành độ $\bar{x} - \sigma_x$ và $\bar{x} + \sigma_x$ là 2 điểm uốn đường cong. Sau đó tung độ giảm chậm và cuối cùng bên trái cũng như bên phải đường cong tiệm cận với trục hoành (xem hình).

b. Các tham số đặc trưng

- Số bình quân:
$$\bar{x} = \int_{-\infty}^{+\infty} xf(x) dx$$

- Phương sai:
$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx$$

- Độ lệch chuẩn:
$$\sigma_x = \sqrt{\sigma_x^2}$$

Số bình quân xác định vị trí đường cong trên trục x, còn độ lệch chuẩn xác định hình dáng

nhọn hay dẹt của đường cong. (Độ lệch chuẩn bé thì đường cong nhọn, độ lệch chuẩn lớn thì đường cong dẹt).

Vì đường cong đối xứng qua \bar{x} nên số bình quân vừa là số mốt vừa là số trung vị. $\bar{x} = M_0 = Me$

Hàm $f(x)$ là hàm mật độ xác suất nên có thể chứng minh được:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

c. Đường cong chuẩn thu gọn

Nếu trong công thức đặt:

$$t = \frac{x - \bar{x}}{\sigma_x}$$

Khi đó số bình quân theo biến mới (t) sẽ là:

$$\bar{t} = \frac{\bar{x} - \bar{x}}{\sigma_x} = 0$$

Các phương sai theo biến t:

$$\sigma_t^2 = \frac{\sigma_{x-\bar{x}}^2}{\sigma_x^2} = \frac{1}{\sigma_x^2} \sigma_{(x)}^2 = 1$$

Như vậy từ cách biến đổi công thức có thể viết thành dạng:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Hàm $\varphi(t)$ được gọi là phương trình đường cong chuẩn thu gọn, còn t được gọi là biến chuẩn hóa. Ở đây t được phân phối theo luật chuẩn với số bình quân $\bar{t} = 0$ và độ lệch chuẩn $\sigma_t = 1$.

2.5.2. So sánh phân phối thực nghiệm với phân phối lý thuyết

1. So sánh bằng tiêu chuẩn λ^2

Khi so sánh phân phối thực nghiệm với một phân phối lý thuyết bằng tiêu chuẩn λ^2 tiến hành theo các bước sau:

Bước 1: Vẽ đồ thị tần số thực nghiệm qua đó phán đoán dạng phân phối lý thuyết.

Do có nhiều dạng phân phối lý thuyết, trong số đó thường chỉ chọn ra 1 dạng để đem so

sánh. Vì vậy để đỡ tốn thời gian phải chọn nhiều lần, ngày từ đầu cần có sự phán đoán tương đối chính xác dạng phân phối lý thuyết. Mỗi dạng phân phối lý thuyết có một đồ thị biểu diễn nhất định. Thông qua đồ thị tần số thực nghiệm sẽ có sự liên hệ và căn cứ hợp lý để phán đoán. Tuy nhiên điều phán đoán được vẫn không thể thay thế được kết luận cuối cùng.

Bước 2: Tính các tham số cần thiết của phân phối thực nghiệm

- Xác suất p, q (đối với phân phối lý thuyết và phân phối nhị thức)
- Số bình quân
- Độ lệch chuẩn
- Xác suất P (x)
- Tần số lý thuyết f_i

Cách xác định các tham số này đã được trình bày cụ thể trong phần trên.

Bước 3: Tính giá trị tiêu chuẩn λ^2 và so sánh kết luận

Giá trị tiêu chuẩn λ^2 được tính theo công thức:

$$\lambda^2 = \sum_{i=1}^k \frac{(f_i - f_i')^2}{f_i'}$$

Trong đó: f_i - Tần số thực nghiệm

f_i' - Tần số lý thuyết

$i = 1, 2, \dots, k$ - Thứ tự các tổ.

Chú ý: Khi áp dụng công thức cần thỏa mãn điều kiện trong tất cả các tổ tần số lý thuyết cũng như tần số thực nghiệm đều phải lớn hơn hoặc bằng 5. Nếu chưa thỏa mãn cần phải tiến hành ghép tổ.

Sau khi tính được λ^2 đem so sánh với giá trị λ_{bang}^2 (tra được từ bảng tính sẵn với độ tự do T và hệ số sai lầm cho phép Ps)

- Nếu $\lambda^2 < \lambda_{bang}^2$ kết luận phân phối thực nghiệm phù hợp với phân phối lý thuyết với hệ số sai lầm Ps.

- Nếu $\lambda^2 \geq \lambda_{bang}^2$ không có cơ sở kết luận phù hợp của 2 phân phối. Lúc đó tạm thời coi phân phối thực nghiệm không phù hợp với phân phối lý thuyết với hệ số sai lầm Ps.

2. So sánh bằng tiêu chuẩn Romanosky

Ở đây các bước tiến hành cũng giống như khi sử dụng tiêu chuẩn λ^2 . Nhưng để

đánh giá kết quả và rút ra kết luận cuối cùng không sử dụng λ_{bang}^2 mà dùng chỉ tiêu R.

$$R = \frac{|\lambda^2 - T|}{\sqrt{2T}}$$

Trong đó: λ^2 - Tính theo công thức

T - Độ tự do

Căn cứ vào giá trị của R để kết luận:

- Nếu $R < 3$ phân phối thực nghiệm phù hợp với phân phối lý thuyết
- Nếu $R \geq 3$ phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

3. So sánh bằng tiêu chuẩn Kolmogorop

Việc so sánh cũng được tiến hành theo ba bước, trong đó bước 1 và bước 2 giống như khi so sánh bằng tiêu chuẩn λ^2 .

Bước 3: Tính giá trị tiêu chuẩn và so sánh. Tiêu chuẩn Kolmogorop D được xác định bằng công thức:

$$D = \frac{1}{n} \max |F_i - F_i'|$$

Trong đó: F_i - Tần số thực nghiệm cộng dồn

F_i' - Tần số lý thuyết cộng dồn

$$n - \text{Tổng các tần số} \quad n = \sum f_i = \sum f_i'$$

Giá trị của D tính theo công thức được so sánh với $D_{bảng}$

- Nếu $D < D_{bảng}$ kết luận với hệ số sai lầm P_s phân phối thực nghiệm phù hợp với phân phối lý thuyết.

- Nếu $D \geq D_{bảng}$ phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

Khi sử dụng tiêu chuẩn Kolmogorop không cần phải ghép các tổ có tần số nhỏ hơn 5.

2.5.3 Các chỉ tiêu biểu thị hình dáng của phân phối

Các chỉ tiêu biểu thị hình dáng của phân phối chủ yếu sử dụng đối với phân phối chuẩn.

1. Các chỉ tiêu biểu hiện sự không đối xứng của phân phối

Để biểu hiện sự không đối xứng của phân phối có thể sử dụng hai cách phổ biến sau:

So sánh trung bình, Mốt và trung vị, cụ thể:

- Nếu: $\bar{x} = M_0 = Me$ dãy số có phân phối chuẩn đối xứng
- Nếu: $\bar{x} > M_0 = Me$ dãy số có phân phối chuẩn lệch phải
- Nếu: $\bar{x} < M_0 < Me$ dãy số có phân phối chuẩn lệch trái

Hệ số không đối xứng:

$$K_A = \frac{\bar{x} - M_0}{\sigma}$$

- $K_A > 0$ dãy số có phân phối chuẩn lệch phải
- $K_A < 0$ dãy số có phân phối chuẩn lệch trái
- K_A càng lớn dãy số có phân phối càng không đối xứng
- $K_A = 0$ dãy số có phân phối chuẩn đối xứng

2. Các chỉ tiêu biểu hiện chiều cao và độ dốc của phân phối

Thường sử dụng các Mômen phân phối:

Công thức tính tổng quát:

$$M_k = \overline{(x_i - A)^k} \text{ gọi là mômen bậc } k$$

Trong đó A là đại lượng không đổi.

- Nếu A = 0:

$$M_k = \overline{(x_i - A)^k} = \frac{\sum x_i^k}{n}$$

Mômen ban đầu bậc 1:

$$M_k = \frac{\sum x_i}{n} = \bar{x}$$

- Nếu A \neq 0 và A = x_0 tùy ý thì:

$$M'_k = \overline{(x_i - x_0)^k} = \frac{\sum (x_i - x_0)^k}{n}$$

- Nếu A = \bar{x} ta có mômen trung tâm:

$$\mu_k = \overline{(x_i - \bar{x})^k} = \frac{\sum (x_i - \bar{x})^k}{n}$$

Khi k = 2 gọi là mômen trung tâm bậc 2:

$$\mu_2 = \overline{(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})^2}{n} = \sigma^2$$

Khi k = 3 gọi là mômen trung tâm bậc 3:

$$\mu_3 = \overline{(x_i - \bar{x})^3} = \frac{\sum(x_i - \bar{x})^3}{n}$$

Khi k = 4 gọi là mômen trung tâm bậc 4:

$$\mu_4 = \overline{(x_i - \bar{x})^4} = \frac{\sum(x_i - \bar{x})^4}{n}$$

Đánh giá tính chất đối xứng của phân phối:

$$A_3 = \frac{\mu_3}{\sigma^3}$$

- $A_3 = 0$ dãy số có phân phối chuẩn đối xứng
- $A_3 > 0$ dãy số có phân phối chuẩn lệch phải
- $A_3 < 0$ dãy số có phân phối chuẩn lệch trái

Biểu hiện độ dốc của phân phối:

$$A_4 = \frac{\mu_4}{\sigma^4}$$

- $A_4 = 0$ dãy số có phân phối có độ dốc gần giống với phân phối chuẩn
- $A_4 > 0$ dãy số có phân phối có độ dốc cao hơn phân phối chuẩn
- $A_4 < 0$ dãy số có phân phối có độ dốc thấp hơn phân phối chuẩn

TÀI LIỆU THAM KHẢO CHƯƠNG 2

1. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế.**

Nhà xuất bản Thống kê, 2004

2. TS. Hồ Sỹ Chi - **Thống kê doanh nghiệp.** NXB Tài chính, 2000

3. GS.TS. Phạm Ngọc Kiểm - **Giáo trình Thống kê doanh nghiệp.** NXB Lao động - Xã hội, 2012.

4. GS.TS. Phạm Ngọc Kiểm TS. Nguyễn Công Nhự - **Giáo trình Thống kê doanh nghiệp.** Nhà xuất bản Thống kê, 2007

5. TS. Nguyễn Công Nhự - **Giáo trình Thống kê công nghiệp.** NXB Thống kê, 2003

6. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2005

7. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2007

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 2

1. Khái niệm số tuyệt đối, phân loại, cho ví dụ minh họa.
2. Khái niệm số tương đối, đặc điểm số tương đối, cho ví dụ minh họa
3. Các loại số tương đối, cho ví dụ minh họa.
4. Khái niệm, ý nghĩa của số bình quân.
5. Thế nào là số bình quân nhân? Cho ví dụ minh họa
6. Thế nào là Một? Phương pháp xác định một? Cho ví dụ minh họa.
7. Thế nào là trung vị? Phương pháp xác định trung vị? Cho ví dụ minh họa.
8. Trình bày các chỉ tiêu đánh giá mức độ biến thiên của tiêu thức. Cho ví dụ minh họa.
9. Quan sát 100 Bưu phẩm của khách hàng gửi tại một Bưu cục có kết quả như sau:

Trọng lượng bưu phẩm (gam)	Số bưu phẩm
<20	2
20-24	12
24-28	14
28-32	15
32-36	20
36-40	10
40-44	12
44-48	3
>48	2
Cộng	100

Hãy khảo sát sự phân phối của trọng lượng bưu phẩm (Sử dụng tiêu chuẩn λ^2)

10. Khảo sát thời gian đàm thoại của 100 khách hàng tại một Bru cục có kết quả như sau:

Thời gian đàm thoại (phút)	Số khách hàng
1	4
2	8
3	14
4	20
5	18
6	13
7	10
8	7
9	2
10	2
11	1
12	1
Cộng	100

Hãy khảo sát sự phân phối thời gian đàm thoại của khách hàng (Sử dụng tiêu chuẩn Romanosky)

11. Quan sát 200 khách hàng vào siêu thị mua hàng ta có số liệu sau:

Số sản phẩm	Số khách hàng
0	18
1	40
2	64

3	40
4	22
5	10
6	6
Cộng	200

Số sản phẩm mà khách hàng mua có tuân theo phân phối nhị thức không ? (Sử dụng tiêu chuẩn Kolmogorop)

CHƯƠNG 3

ĐIỀU TRA CHỌN MẪU

3.1. Khái niệm điều tra chọn mẫu

Trong phần trước đã trình bày các loại điều tra thống kê, nếu theo số lượng các đơn vị của hiện tượng điều tra thực tế thì có thể phân các cuộc điều tra thống kê thành hai loại là điều tra toàn bộ và điều tra không toàn bộ.

Điều tra chọn mẫu là một loại điều tra không toàn bộ trong đó người ta chỉ chọn ra một số đơn vị trong toàn bộ các đơn vị của hiện tượng nghiên cứu để điều tra thực tế, rồi dùng các kết quả tính toán để suy rộng thành các đặc điểm của toàn bộ tổng thể nghiên cứu.

Ví dụ: Để đánh giá về đời sống nhân dân của một địa phương nào đó thì có thể chọn ra một số hộ để thu thập tài liệu về lao động, về nghề nghiệp, về tình hình thu chi,.. Dựa vào tài liệu điều tra để suy rộng về đời sống của nhân dân toàn bộ địa phương đó.

Tại sao chỉ điều tra 1 số đơn vị tổng thể mà suy ra kết quả cho cả tổng thể. Vì quy luật số lớn đã chỉ ra rằng nếu nghiên cứu 1 số tương đối lớn hiện tượng thì những biểu hiện ngẫu nhiên, những đặc thù của hiện tượng riêng lẻ sẽ bù trừ và triệt tiêu cho nhau, tính quy luật sẽ được biểu hiện rõ. Hơn nữa lý thuyết xác suất cũng chứng minh rằng sự sai khác giữa số bình quân của một số rất lớn đại lượng ngẫu nhiên với kỳ vọng toán của nó là một đại lượng nhỏ tùy ý.

Khi chọn đơn vị để điều tra chọn mẫu người ta có thể chọn theo cách ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau không phụ thuộc vào ý kiến chủ quan người chọn mẫu, hoặc không ngẫu nhiên (bàn bạc, phân tích tập thể để lựa chọn ra những đơn vị đại biểu).

So với điều tra toàn bộ điều tra chọn mẫu có những ưu điểm chủ yếu sau:

Thứ nhất, điều tra chọn mẫu thường nhanh hơn rất nhiều so điều tra toàn bộ, vì điều tra ít đơn vị, nên công tác chuẩn bị nhanh gọn, số lượng tài liệu ghi chép ít, thời gian điều tra ghi chép, tổng hợp và phân tích sẽ được rút ngắn. Do đó điều tra chọn mẫu mang tính kịp thời cao.

Thứ hai, do điều tra ít đơn vị, số nhân viên điều tra và mọi chi phí điều tra giảm, cho nên điều tra chọn mẫu sẽ tiết kiệm sức người, vật tư, tiền của.

Thứ ba, do điều tra ít đơn vị nên có thể mở rộng được nội dung điều tra, đi sâu nghiên cứu nhiều mặt của hiện tượng.

Thứ tư, tài liệu thu thập được trong điều tra chọn mẫu sẽ có độ chính xác cao, bởi vì số nhân viên điều tra ít nên có thể chọn được những người có trình độ chuyên môn cao và nhiều kinh nghiệm, đồng thời việc kiểm tra số liệu có thể tiến hành tỷ mỉ và tập trung, do đó các sai sót do ghi chép sẽ giảm đi nhiều.

Thứ năm, điều tra chọn mẫu không đòi hỏi một tổ chức lớn như điều tra toàn bộ, một tổ chức nhỏ cũng có thể tiến hành điều tra chọn mẫu.

Tóm lại điều tra chọn mẫu có rất nhiều ưu điểm so với điều tra toàn bộ, nhưng do điều tra chọn mẫu tiến hành với phạm vi nhỏ nên sẽ có sai số nhất định so với kết quả điều tra toàn bộ nên không thể dùng điều tra chọn mẫu để hoàn toàn thay thế cho điều tra toàn bộ.

Điều tra chọn mẫu có thể sử dụng trong các trường hợp sau:

- Dùng để thay thế điều tra toàn bộ khi đối tượng nghiên cứu cho phép vừa có thể điều tra toàn bộ, vừa có thể điều tra chọn mẫu, thì tiến hành điều tra chọn mẫu để có kết quả nhanh và tiết kiệm hoặc dùng cho những trường hợp việc điều tra có liên quan tới phá huỷ đơn vị điều tra.

- Kết hợp với điều tra toàn bộ để mở rộng nội dung điều tra và đánh giá kết quả điều tra toàn bộ.

- Dùng trong trường hợp muốn so sánh các hiện tượng với nhau mà chưa có tài liệu cụ thể, hoặc kiểm định giả thiết đặt ra.

- Dùng để tổng hợp nhanh tài liệu điều tra toàn bộ, có thông tin nhanh, phục vụ kịp thời cho công tác quản lý.

Phân loại điều tra chọn mẫu

Điều tra chọn mẫu ngẫu nhiên

Điều tra chọn mẫu ngẫu nhiên là phương pháp chọn mẫu trong đó các đơn vị tổng thể được chọn vào mẫu một cách hoàn toàn ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau, không phụ thuộc vào ý muốn chủ quan người chọn mẫu.

Điều tra chọn mẫu phi ngẫu nhiên.

Điều tra chọn mẫu phi ngẫu nhiên là sự lựa chọn các đơn vị vào mẫu điều tra dựa trên kinh nghiệm và hiểu biết của người nghiên cứu về tổng thể nghiên cứu, hoặc căn cứ vào những qui định nhất định khi lấy mẫu.

3.2. Điều tra chọn mẫu ngẫu nhiên

3.2.1 Tổng thể chung và tổng thể mẫu

Từ khái niệm về điều tra chọn mẫu cho thấy có hai tổng thể là tổng thể chung và tổng thể mẫu.

Tổng thể chung là tổng thể bao gồm toàn bộ các đơn vị thuộc đối tượng điều tra. Số đơn vị của tổng thể chung thường được ký hiệu là N và được biểu thị bằng dãy số lượng biến.

$$X_1, X_2, X_3, \dots, X_n, \dots, X_N$$

Tổng thể mẫu là tổng thể bao gồm một số đơn vị nhất định được chọn ra từ tổng thể chung để điều tra thực tế. Số đơn vị của tổng thể mẫu thường được ký hiệu là n . Tổng thể mẫu gồm n đơn vị tổng thể thường được biểu diễn dưới dạng bản phân phối.

Có hai cách chọn ngẫu nhiên số lượng n đơn vị của tổng thể mẫu:

Thứ nhất, chọn mẫu nhiều lần (chọn hoàn lại)

Chọn mẫu nhiều lần là khi mỗi đơn vị được chọn ra để đăng ký rồi lại trả về tổng thể chung. Như vậy số đơn vị tổng thể chung không thay đổi trong suốt quá trình lấy mẫu. Xác suất được chọn của mỗi đơn vị tổng thể là bằng nhau.

Gọi K là số khả năng thiết lập được tổng thể mẫu thì số khả năng đó trong chọn nhiều lần được xác định theo công thức sau:

$$K = N^n$$

Thứ hai, chọn mẫu 1 lần (chọn không hoàn lại)

Chọn mẫu một lần là khi mỗi đơn vị đã được chọn để đăng ký rồi sẽ được xếp riêng ra không trả lại về tổng thể chung nữa, do đó không có khả năng chọn lại, xác suất được chọn của mỗi đơn vị không bằng nhau.

Khi đó số khả năng thiết lập tổng thể mẫu được xác định theo công thức:

$$K = C_n^N = \frac{N!}{n!(N-n)!}$$

Như vậy với cả hai cách chọn mẫu ngẫu nhiên thì số lượng mẫu có thể hình thành là rất lớn. Mẫu được chọn ra để điều tra chỉ là một trong số rất lớn số lượng mẫu có thể được hình thành.

Nếu ký hiệu:

- \bar{X}, P, σ^2 - là bình quân, tỉ lệ, phương sai của tổng thể chung.

- $\bar{x}_i, w_i, \sigma_i^2$ - là bình quân, tỉ lệ, phương sai của tổng thể mẫu thứ i ($i=1,2,3,\dots,Q$) và thực chất đó là biến ngẫu nhiên và người ta đã chứng minh được rằng: Nếu $n \geq 30$ thì \bar{x}_i, w_i được xem như phân phối chuẩn với:

- Kỳ vọng các bình quân mẫu:

$$E(\bar{x}_i) = \bar{X}$$

- Phương sai các bình quân mẫu:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad (\text{chọn hoàn lại})$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right) \quad (\text{chọn không hoàn lại})$$

- Kỳ vọng các tỉ lệ mẫu:

$$E(w_i) = p$$

- Phương sai các tỉ lệ mẫu:

$$\sigma_w^2 = \frac{p(1-p)}{n} \quad (\text{chọn hoàn lại})$$

$$\sigma_w^2 = \frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right) \quad (\text{chọn không hoàn lại})$$

Nếu $n < 30$ thì \bar{x}_l được xem phân phối theo quy luật Student

3.2.2 Chọn mẫu với xác suất đều và xác suất không đều

Chọn mẫu với xác suất đều là đảm bảo mỗi đơn vị của hiện tượng nghiên cứu đều có cơ hội được chọn vào mẫu như nhau. Tính bình đẳng còn thể hiện trong việc ước lượng kết quả vì kết quả thu được trên mẫu không phân biệt đơn vị nào. Phương pháp chọn mẫu không đều không lưu ý đến sự khác biệt giữa các đơn vị tổng thể và thường được sử dụng khi các đơn vị của tổng thể tương đối đồng đều nhau theo tiêu thức nghiên cứu, hoặc không biết trước được sự khác biệt giữa các đơn vị điều tra.

Chọn mẫu với xác suất không đều nghĩa là không cần đảm bảo khả năng được chọn vào mẫu của các đơn vị phải bằng nhau. Các đơn vị có thể được chọn vào mẫu theo xác suất tỷ lệ vai trò của từng đơn vị. xác suất ấn định riêng cho mỗi đơn vị về khả năng được chọn vào mẫu là xác suất bao hàm.

Việc chọn mẫu với xác suất không đều có khó khăn, phức tạp hơn vì phải có số liệu tiêu nghiệm về qui mô, kích cỡ của từng đơn vị.

3.2.3. Sai số trong chọn mẫu

Sai số chọn mẫu là sự chênh lệch về trị số giữa các chỉ tiêu tính ra được trong điều tra chọn mẫu và các chỉ tiêu tương ứng của tổng thể chung.

Các loại sai số

a. Sai số chọn mẫu:

Như đã nói, ta không biết được một cách chính xác các đặc trưng của tổng thể như trung bình hoặc tỷ lệ, mà chỉ có thể ước lượng chúng dựa trên thông tin thu thập từ mẫu. Các ước lượng đó, tất nhiên là không thể tránh khỏi sai sót, gọi là sai số chọn mẫu. Như vậy sai số chọn mẫu là sai số do sử dụng thông tin thu thập được chỉ trên một bộ phận tổng thể, hoặc là do mẫu không thể hiện, đại diện được cho cả tổng thể. Nói cách khác sai số chọn mẫu do hai yếu tố: yếu tố thiết kế, thực hiện và yếu tố ngẫu nhiên. Có thể hạn chế sai số bằng cách hết sức chú ý tới việc thiết kế, thực hiện chọn mẫu, song sai số chọn mẫu hầu như là tất nhiên, chỉ có thể giảm bớt bằng cách tăng qui mô của mẫu.

b. Sai số phi chọn mẫu:

Sai số phi chọn mẫu là sai số không thuộc về phương pháp chọn mẫu được sử dụng. Các điều kiện để phát sinh sai số phi chọn mẫu có thể là:

- Sai số do ghi chép: do đơn vị điều tra chưa hiểu đúng nội dung câu hỏi nên trả lời sai, do đo lường, do vô tình ghi chép sai hoặc cố ý ghi chép sai do mục đích nào đó.

Nếu là sai số ngẫu nhiên thì các sai lệch trong khi lấy số liệu có thể bù trừ nhau. Loại sai số này ít gây nguy hiểm.

Sai số hệ thống là sai số nguy hiểm, càng nhiều đơn vị điều tra thì sai số càng nhiều. Nó xảy ra do dụng cụ đo lường sai hay cố ý ghi sai.

Để có thể giảm bớt được sai số do ghi chép cần phải chuẩn bị tốt cho công tác điều tra, trình độ dụng cụ máy móc và ý thức người điều tra.

- Sai số do mẫu được lấy từ một tổng thể không thích hợp.

- Sai số do tỷ lệ không trả lời quá cao. Việc không nhận được các câu trả lời là một vấn đề quan trọng phải giải quyết của phương pháp chọn mẫu. Nó làm cho kết quả điều tra bị sai lệch do thông tin không đầy đủ, tổng thể trả lời có thể rất khác xa với tổng thể thực sự muốn điều tra.

3.2.4. Sai số bình quân chọn mẫu và phạm vi sai số chọn mẫu

1. Sai số bình quân chọn mẫu

a. Khi điều tra chọn mẫu dùng để suy rộng số trung bình của một tiêu thức nào đó thì sai số được tính như sau:

+ Khi chọn mẫu nhiều lần

$$\mu_x = \sqrt{\frac{\sigma^2}{n}}$$

Trong đó: μ_x - Sai số trung bình chọn mẫu.

σ^2 - Phương sai chung.

n - Số đơn vị tổng thể mẫu.

Do phương sai tổng thể chung không tính được nên phải dùng phương sai mẫu điều chỉnh để thay thế:

$$\mu_x = \sqrt{\frac{\sigma_n^2}{n-1}}$$

+ Khi chọn mẫu 1 lần

$$\mu_x = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

$$\text{hay } \mu_x = \sqrt{\frac{\sigma_n^2}{n-1} \left(1 - \frac{n}{N}\right)}$$

Trong đó: σ^2_n - là phương sai mẫu điều chỉnh:

N - Số đơn vị của tổng thể chung.

Ví dụ : Tính sai số chọn mẫu ngẫu nhiên nhiều lần trong trường hợp tính năng suất lao động (NSLĐ) chung của doanh nghiệp A có tổng số công nhân là 200 người, trong đó chọn ra 100 công nhân để điều tra thực tế về kết quả điều tra như sau:

Bảng 3.1

NSLĐ của 1 CN x_i	Số công nhân f_i	x'_i	$f_i x'_i$	$x'_i - \bar{x}$	$(x'_i - \bar{x})^2$	$f_i (x'_i - \bar{x})^2$
35-45	14	20	560	-18	324	4536
45-55	20	50	1000	-8	64	1280
55-65	42	60	2520	2	4	168
65-75	20	70	1400	12	144	2880
75-85	4	80	320	22	484	1936
Cộng	100		5800			10800

Ta có:

$$\bar{x} = \frac{\sum x'_i f_i}{\sum f_i} = \frac{5800}{100} = 58(T)$$

$$\sigma_m^2 = \frac{10800}{100} = 108$$

$$\mu_x = \sqrt{\frac{108}{100-1}} = 1,044$$

b. Khi điều tra chọn mẫu để suy rộng tỷ lệ theo một tiêu thức nào đó thì sai số được tính như sau:

- Trường hợp chọn mẫu nhiều lần:

$$\mu_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Trong đó: p - Tỷ lệ cấu thành của tổng thể chung:

$$p = \frac{M}{N}$$

M - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm, theo định nghĩa xác suất thì p chính là xác suất để lấy đơn vị mang dấu hiệu A trong tổng thể chung.

q - là tỷ trọng của bộ phận không mang dấu hiệu A của tổng thể chung.

Do tỷ lệ cấu thành chung không tính được nên phải dùng tỷ lệ cấu thành mẫu điều chỉnh để thay thế:

$$\mu_p = \sqrt{\frac{w(1-w)}{n-1}}$$

Trong đó: w - Tỷ lệ cấu thành của tổng thể mẫu $w = \frac{m}{n}$

m - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm thuộc tổng thể mẫu.

n - Số đơn vị của tổng thể mẫu.

- Trong trường hợp chọn mẫu 1 lần

$$\mu_p = \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

$$\mu_p = \sqrt{\frac{w(1-w)}{n}}$$

Ý nghĩa của μ_x và μ_p

Sai số μ_x và μ_p biểu hiện trị số của các sai số chọn mẫu gặp phải khi suy rộng tài liệu. Nhưng do chọn mẫu ngẫu nhiên nên sai số này không được xác định trước về dấu(+ hoặc -) mà nó phản ánh 1 phạm vi chênh lệch có thể nhiều hơn hay ít hơn so với tham số của tổng thể chung, do đó chênh lệch giữa số trung bình mẫu (tỷ lệ mẫu) và số trung bình chung (tỷ lệ chung) là $\pm \mu_x$ và μ_p , nghĩa là:

$$\bar{X} = \bar{x} \pm \mu_x$$

$$p = w \pm \mu_p$$

\bar{X} - là số trung bình chung;

\bar{x} - Số trung bình mẫu.

p - Tỷ lệ chung.

w - Tỷ lệ mẫu.

2. Phạm vi sai số chọn mẫu

Phạm vi sai số chọn mẫu có thể được xác định theo công thức:

$$\varepsilon = t\mu$$

Trong đó: ε - Phạm vi sai số chọn mẫu.

t - Hệ số tin cậy

μ - sai số trung bình chọn mẫu.

Hoặc

$$\varepsilon_x = t\mu_x$$

$$\varepsilon_p = t\mu_p$$

3. Suy rộng kết quả điều tra chọn mẫu

Tức là tính toán các tham số của tổng thể chung trên cơ sở thu thập được trong điều tra chọn mẫu. Có hai phương pháp suy rộng:

a. *Phương pháp tính đối trực tiếp:*

Được áp dụng khi dùng các số trung bình mẫu hay tỷ lệ mẫu của tổng thể mẫu để tính ra các tham số của tổng thể chung:

$$\begin{aligned}\bar{X} &= \bar{x} \pm \varepsilon_x \Rightarrow \\ \bar{x} - \varepsilon_x &\leq \bar{X} \leq \bar{x} + \varepsilon_x\end{aligned}$$

$$p = w \pm \varepsilon_p$$

$$w - \varepsilon_p \leq p \leq w + \varepsilon_p$$

b. *Phương pháp hệ số điều chỉnh:*

Phương pháp hệ số điều chỉnh được dùng để xác định kết quả điều tra toàn bộ. Căn cứ vào kết quả điều tra toàn bộ và điều tra chọn mẫu, tính ra tỷ lệ chênh lệch rồi dùng tỷ lệ này để làm hệ số điều chỉnh kết quả điều tra toàn bộ.

4. Xác định số đơn vị tổng thể mẫu

Khi điều tra chọn mẫu cần xác định số đơn vị điều tra sao cho thoả mãn hai yêu cầu sau:

- Bảo đảm sai số chọn mẫu nhỏ nhất;
- Chi phí thấp nhất.

Hai yêu cầu này đối lập nhau. Muốn đảm bảo tài liệu điều tra chính xác (sai số nhỏ nhất) phải điều tra một lượng đơn vị khá lớn, tức là cần một lượng kinh phí lớn. Ngược lại với một lượng kinh phí có hạn, chỉ có thể điều tra một số ít đơn vị thì phải chấp nhận một sai số nhất định.

Trong thực tế thường căn cứ vào yêu cầu của độ chính xác (phạm vi sai số chọn mẫu) để tính số đơn vị mẫu cần điều tra.

+ Trường hợp để ước lượng số trung bình:

- Nếu dự định chọn đơn vị tổng thể mẫu theo cách chọn nhiều lần (chọn hoàn lại):

$$\varepsilon_x = t\mu_x = t\sqrt{\frac{\delta^2}{n}} \quad \text{suy ra } n = \frac{t^2\delta^2}{\varepsilon_x^2}$$

- Trong trường hợp chọn một lần (chọn không hoàn lại):

$$\varepsilon_x = t\sqrt{\frac{\delta_x^2}{n}\left(1 - \frac{n}{N}\right)} \quad \text{suy ra } n = \frac{Nt^2\delta^2}{N\varepsilon_x^2 + t^2\delta^2}$$

+ Trường hợp để ước lượng tỷ lệ:

- Nếu dự định chọn các đơn vị tổng thể theo cách chọn nhiều lần:

$$\varepsilon_p = t\mu_p = t\sqrt{\frac{pq}{n}} \quad \text{suy ra } n = \frac{t^2pq}{\varepsilon_p^2}$$

- Nếu dự định chọn mẫu theo cách chọn một lần:

$$\varepsilon_p = t\mu_p = t\sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)} \quad \text{suy ra } n = \frac{Nt^2pq}{N\varepsilon_p^2 + t^2pq}$$

Qua các công thức cho thấy các nhân tố ảnh hưởng đến số đơn vị mẫu điều tra bao gồm:

- Phạm vi sai số chọn mẫu. Nếu yêu cầu phạm vi sai số càng nhỏ thì số đơn vị mẫu cần điều tra càng nhiều và ngược lại. Số đơn vị mẫu tỷ lệ nghịch với bình phương phạm vi sai số chọn mẫu.

- Hệ số tin cậy (t). Yêu cầu trình độ tin cậy của tài liệu suy rộng cao hay thấp thì hệ số tin cậy cũng phải lớn hay nhỏ một cách tương ứng và do đó số đơn vị mẫu cũng phải tăng giảm cho phù hợp. Số đơn vị mẫu tỷ lệ thuận với bình phương hệ số tin cậy.

- Tính chất đồng đều của hiện tượng nghiên cứu. Số đơn vị mẫu tỷ lệ thuận với phương sai của tổng thể chung σ^2 và pq. Nếu tiêu thức biến thiên ít, phương sai tính ra nhỏ, số đơn vị cần điều tra sẽ giảm đi. Ngược lại nếu tiêu thức biến thiên lớn, phương sai sẽ lớn làm cho số đơn vị cần điều tra sẽ tăng lên.

Trong thực tế, khi tính số đơn vị mẫu cần điều tra thường không có tài liệu để tính phương sai chung, vì vậy phải giải quyết bằng cách lấy phương sai lớn nhất của các lần điều tra trước hoặc lấy phương sai của hiện tượng tương tự, hoặc tổ chức điều tra thí điểm để lấy phương sai.

Sau khi xác định được số đơn vị mẫu điều tra, người ta thường kiểm tra tính chất đại biểu của tổng thể đó. Đây là việc làm cần thiết vì nếu tổng thể mẫu không đại biểu cho tổng thể chung thì kết quả suy rộng sẽ không chính xác. Việc kiểm tra tính chất đại biểu của tổng thể mẫu có thể tiến hành bằng nhiều cách:

So sánh một số chỉ tiêu trung bình hoặc chỉ tiêu tỷ lệ của tổng thể mẫu đã chọn với các chỉ tiêu tương ứng của tổng thể chung (\bar{X} với \bar{x} , w với p). Tất nhiên là giữa những chỉ tiêu này thường có sự chênh lệch, nhưng nếu mức độ chênh lệch không vượt quá phạm vi cho phép ($\pm 5\%$) thì tổng thể mẫu được coi như đủ tính chất đại biểu. nếu mức độ chênh lệch vượt quá phạm vi $\pm 5\%$, tổng thể mẫu chưa đủ tính chất đại biểu, do đó phải thay thế một số đơn vị hoặc tăng thêm số đơn vị điều tra.

Kiểm tra và sử lý những số liệu nghi ngờ. Nếu trong quá trình chọn các đơn vị điều tra gặp phải một vài đơn vị có số liệu quá lớn hay quá bé thì cần kiểm tra và sử lý bằng cách:

+ Nếu số đơn vị có số liệu đột xuất không nhiều, thì nên loại bỏ thay thế bằng các đơn vị khác.

+ Nếu số đơn vị có số liệu đột xuất không phải là cá biệt, thì nên chọn phân loại để đảm bảo có một số mẫu nhất định đại diện cho tổng thể chung.

+ Có thể so sánh sự phân phối các đơn vị theo tiêu thức nghiên cứu của 2 tổng thể. Cách so sánh này phức tạp hơn nhưng đảm bảo tính đại biểu hơn.

3.3. Quy trình một cuộc điều tra chọn mẫu ngẫu nhiên

Một cuộc điều tra chọn mẫu ngẫu nhiên thường được tiến hành qua các giai đoạn sau đây:

3.3.1. Xác định mục tiêu nghiên cứu

Xác định mục tiêu nghiên cứu tức là phải xác định một cách rõ ràng là cuộc điều tra đó nhằm tìm hiểu những vấn đề gì, phục vụ cho yêu cầu cụ thể nào?

Xác định mục tiêu nghiên cứu là bước khởi đầu rất quan trọng, là tiền đề cho giai đoạn sau.

3.3.2. Xác định tổng thể nghiên cứu

Tổng thể nghiên cứu là tổng thể chung bao gồm tất cả các đơn vị của hiện tượng nghiên cứu tức xác định N. Để xác định tổng thể nghiên cứu thì phải dựa vào mục đích nghiên cứu, mục đích nghiên cứu khác nhau thì tổng thể nghiên cứu khác nhau. Thực chất của việc xác định tổng thể nghiên cứu là đi xác định các đơn vị của nó. Muốn vậy, ngoài việc dựa vào mục đích nghiên cứu còn phải dựa vào lý luận kinh tế xã hội, tình hình thực tế để đưa ra định nghĩa, những tiêu chuẩn làm căn cứ để xem xét một cách cụ thể.

3.3.3. Xác định nội dung điều tra

Xác định nội dung điều tra là xác định danh mục các tiêu thức cần điều tra trên các đơn vị của tổng thể mẫu và được cụ thể hóa bằng phiếu điều tra. Để xác định nội dung điều tra thì phải dựa vào mục đích nghiên cứu. Mục đích nghiên cứu đòi hỏi phải giải quyết nhiều vấn đề thì nội dung điều tra phải bao gồm nhiều tiêu thức.

3.3.4. Xác định số đơn vị của tổng thể mẫu và phương pháp tổ chức chọn mẫu

Xác định số đơn vị của tổng thể mẫu tức là xác định n . Để xác định số đơn vị của tổng thể mẫu cần phải cho trước phạm vi sai số chọn mẫu và xác suất suy rộng tài liệu. Các công thức tính số lượng đơn vị của tổng thể mẫu đã được trình bày ở trên.

Sau khi đã xác định số đơn vị của tổng thể mẫu, dựa vào các đặc điểm của hiện tượng nghiên cứu và khả năng tổ chức điều tra để áp dụng các phương pháp tổ chức chọn mẫu đã được đề cập ở phần trên nhằm xác định các đơn vị cụ thể của tổng thể mẫu.

3.3.5 Tiến hành thu thập tài liệu ở các đơn vị của tổng thể mẫu

Dựa vào phiếu điều tra để tiến hành thu thập tài liệu ở các đơn vị của tổng thể mẫu. Có nhiều phương pháp thu thập tài liệu như: phương pháp đăng ký trực tiếp, phương pháp phỏng vấn trực tiếp,... Tùy thuộc vào điều kiện và tính chất của cuộc điều tra để áp dụng phương pháp thu thập tài liệu cho phù hợp.

3.3.6. Suy rộng kết quả điều tra chọn mẫu

Sau khi đã thu thập được đầy đủ tài liệu ở các đơn vị của tổng thể mẫu, căn cứ vào đó để tiến hành tính toán và suy rộng ra các đặc điểm của tổng thể chung. Có hai phương pháp suy rộng là suy rộng trực tiếp và suy rộng khoảng.

Suy rộng trực tiếp là coi các mức độ của tổng thể mẫu cũng là các mức độ của tổng thể chung. Suy rộng khoảng cách là các mức độ của tổng thể chung được xác định nhận giá trị trong một khoảng nào đó với xác suất cho trước.

3.3.7. Đưa ra kết luận về tổng thể chung

Đây là giai đoạn cuối cùng thể hiện kết quả của quá trình nghiên cứu. Câu hỏi được đặt ra là: "có thể có kết luận gì về tổng thể chung", các kết luận có đáp ứng được mục đích nghiên cứu đã được đặt ra hay không? Có phản ánh được đặc điểm, bản chất của hiện tượng hay không. Từ đó cần đưa ra những giải pháp kiến nghị cụ thể để thúc đẩy sự phát triển của hiện tượng.

Các kết quả nghiên cứu đạt được cần được trình bày thông qua các bảng thống kê, các đồ thị thống kê và báo cáo phân tích.

3.4. Điều tra chọn mẫu phi ngẫu nhiên

Điều tra chọn mẫu phi ngẫu nhiên là chọn có dụng ý trước, nghĩa là dựa trên sự hiểu biết về hiện tượng nghiên cứu, tiến hành bàn bạc, phân tích để lựa chọn những đơn vị điều hình có khả năng đại diện cho tổng thể nghiên cứu để điều tra. Kết quả điều tra thường được dùng để suy

rộng cho toàn bộ tổng thể hoặc để đánh giá một cách tổng quát. Điều tra chọn mẫu phi ngẫu nhiên được sử dụng rộng rãi để nghiên cứu các hiện tượng kinh tế phức tạp, phân tán, không ổn định đòi hỏi phải quan sát, phân tích tỷ mỉ trước khi thu thập tài liệu. Điều tra chọn mẫu phi ngẫu nhiên không hoàn toàn dựa trên cơ sở toán học như chọn mẫu ngẫu nhiên mà chủ yếu đòi hỏi phải kết hợp chặt chẽ giữa phân tích lý luận và thực tế. Sự nhận xét chủ quan của người tổ chức có ảnh hưởng đến chất lượng điều tra. Muốn đảm bảo chất lượng tài liệu điều tra phải giải quyết các vấn đề sau:

3.4.1. Đảm bảo phân tổ chính xác đối tượng điều tra

Vì mỗi đơn vị được chọn ra dù có đầy đủ tính chất đại biểu đến mấy cũng chỉ có khả năng đại diện cho mỗi bộ phận, một loại hình nào đó trong tổng thể phức tạp. Nếu tập hợp được các điển hình của nhiều bộ phận thì chúng có khả năng đại diện cho cả tổng thể hiện tượng phức tạp. Mặt khác, việc phân tổ có tác dụng thu hẹp độ biến thiên tiêu thức trong mỗi bộ phận làm cho việc suy rộng tài liệu càng tỷ mỉ và chính xác. Tổng thể càng phức tạp, việc phân tích càng phải thận trọng, nhiều khi phân tổ còn phải trải qua nhiều bước để có những tổ chi tiết hơn.

3.4.2. Chọn đơn vị điều tra

Trong điều tra chọn mẫu phi ngẫu nhiên, người ta chọn các đơn vị điển hình có khả năng đại diện cho từng bộ phận khác nhau của tổng thể nghiên cứu. Có nhiều cách chọn:

+ Chọn những đơn vị có mức độ tiêu thức gần với số trung bình của từng bộ phận nhất, đồng thời cũng là mức độ phổ biến nhất trong bộ phận đó. Khi chọn phải thông qua quan sát bàn bạc phân tích tập thể thì mới chọn được những đơn vị điều tra có tính đại biểu cao. Sau khi chọn được các đơn vị điều tra phải kiểm tra tính chất đại biểu của các đơn vị đó, nếu chấp nhận được thì mới tiến hành điều tra thực tế.

+ Chọn những đơn vị có kinh nghiệm về một mặt nào đó.

+ Chọn một số địa phương đại diện cho từng vùng kinh tế. Trong các vùng này lại chọn ngẫu nhiên một số đơn vị điều tra.

3.4.3. Xác định số đơn vị điều tra

Điều tra chọn mẫu phi ngẫu nhiên cũng phải dựa trên cơ sở định luật số lớn, nghĩa là cần chọn ra một số đơn vị điều tra nhiều tới mức đủ khả năng đại biểu cho cả tổng thể. Vì chọn phi ngẫu nhiên nên không thể dùng công thức toán học để tính. Muốn xác định số mẫu cần:

+ Căn cứ vào tính chất phức tạp của tổng thể điều tra. Tổng thể càng phức tạp càng cần điều tra nhiều đơn vị.

+ Căn cứ vào kinh nghiệm của các địa phương khác, nước khác, các lần điều tra trước để quyết định số đơn vị cần điều tra thực tế.

+ Căn cứ vào mức độ đòi hỏi của việc nghiên cứu, lực lượng cán bộ và khả năng vật chất để quyết định tăng thêm hay giảm bớt số đơn vị cần điều tra.

3.4.4. Sai số chọn mẫu

Trong điều tra chọn mẫu phi ngẫu nhiên, sai số chọn mẫu không thể tính được bằng công thức toán học, mà phải thông qua nhận xét, so sánh để ước lượng ra. Nếu thấy sai số không lớn lắm có thể chấp nhận được thì dùng kết quả của điều tra chọn mẫu suy rộng ra các tham số của tổng thể chung. Nếu thấy nghi ngờ thì có thể chọn lại và điều tra lại. Ngoài ra có thể dùng phương pháp kiểm định thống kê để xác định chất lượng kết quả điều tra chọn mẫu. Khi suy rộng thì suy rộng trực tiếp, chứ không suy rộng có phạm vi như trong điều tra chọn mẫu ngẫu nhiên. Các đơn vị điều tra được lựa chọn từ các bộ phận khác nhau nên khi suy rộng phải theo thứ tự từng bước và chú ý đến tỷ trọng của mỗi bộ phận trong tổng thể.

3.4.5. Lực lượng tham gia điều tra

Điều tra chọn mẫu phi ngẫu nhiên đòi hỏi phải giải quyết nhiều vấn đề phức tạp, mà cơ sở chủ yếu là dựa trên sự phân tích sâu sắc đối tượng nghiên cứu. Muốn làm tốt công tác điều tra, phải thành thạo về nghiệp vụ, am hiểu hiện tượng nghiên cứu, làm tốt công tác tổ chức vận động để hiểu rõ mục đích nghiên cứu, giải quyết tốt những vấn đề ảnh hưởng tới chất lượng điều tra.

TÀI LIỆU THAM KHẢO CHƯƠNG 3

1. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – Giáo trình **Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
2. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
3. GS.TS Bùi Xuân Phong – Bài giảng **Thống kê doanh nghiệp** (Đào tạo từ xa) – **HV công nghệ BCVT 2007**
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. TS. Hồ Sỹ Chi - **Thống kê doanh nghiệp**. NXB Tài chính, 2000
5. GS.TS. Phạm Ngọc Kiểm - **Giáo trình Thống kê doanh nghiệp**. NXB Lao động - Xã hội, 2012.
6. GS.TS. Phạm Ngọc Kiểm TS. Nguyễn Công Nhự - **Giáo trình Thống kê doanh nghiệp**. Nhà xuất bản Thống kê, 2007

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 3

1. Khái niệm, ưu nhược điểm của điều tra chọn mẫu.
2. Phân loại điều tra chọn mẫu, phạm vi áp dụng của điều tra chọn mẫu.
3. Khái niệm sai số trong điều tra chọn mẫu, nhân tố ảnh hưởng đến sai số trong điều tra chọn mẫu.
4. Trình bày các phương pháp tổ chức chọn mẫu
5. Trình bày một quy trình một cuộc điều tra chọn mẫu ngẫu nhiên.
6. Khái niệm tiêu thức nguyên nhân, tiêu thức kết quả, cho ví dụ minh họa.

7. Trình bày đối tượng áp dụng điều tra chọn mẫu và những ưu nhược điểm của điều tra chọn mẫu.

8. Thế nào là điều tra chọn mẫu ngẫu nhiên, phi mẫu nhiên?

9. Tại sao nếu tăng độ tin cậy khi suy rộng kết quả điều tra thì giá trị suy rộng của kết quả lại giảm?

10. Một doanh nghiệp trồng 1500 ha cafe. Sắp đến vụ thu hoạch người ta chọn ngẫu nhiên 100 điểm (mỗi điểm $1m^2$) để kiểm tra thu được kết quả sau:

Năng suất (Kg)	Số điểm gặt
Dưới 0,40	15
0,40 - 0,45	25
0,45 - 0,50	30
0,50 - 0,55	20
Từ 0,55 trở lên	10

- Tính năng suất lao động bình quân 1ha của toàn doanh nghiệp với xác suất bằng 0,9545.
- Từ câu a, xác định lượng cafe của toàn doanh nghiệp.
- Tỉ lệ diện tích cafe đạt 5 tấn/ha trở lên với xác suất bằng 0,6826.

CHƯƠNG 4

TƯƠNG QUAN VÀ HỒI QUY

4.1. Khái niệm

4.1.1. Tiêu thức nguyên nhân và tiêu thức kết quả

Giữa các hiện tượng kinh tế xã hội tồn tại những mối liên quan ràng buộc lẫn nhau. Ngay trong cùng một hiện tượng nghiên cứu bao gồm nhiều tiêu thức khác nhau, thì những tiêu thức này cũng có những mối liên hệ qua lại nhất định. Nếu xét về mức độ phụ thuộc giữa các tiêu thức thì người ta thường chia mối liên hệ giữa các hiện tượng thành các loại sau:

Liên hệ hàm số: là loại liên hệ hết sức chặt chẽ giữa hai đại lượng nghiên cứu và biểu hiện sự liên hệ này theo những tỷ lệ tương ứng chặt chẽ, tức là khi hiện tượng thay đổi thì nó hoàn toàn quyết định sự thay đổi của hiện tượng liên quan theo một tỷ lệ tương ứng. Trường hợp này ít gặp trong thực tế, thường gặp trong toán học và vật lý.

Liên hệ tương quan: là mối liên hệ không hoàn toàn chặt chẽ như liên hệ hàm số, tức là khi trị số của tiêu thức nguyên nhân thay đổi nó dẫn đến tiêu thức kết quả thay đổi theo. Nhưng sự thay đổi của tiêu thức kết quả không hoàn toàn phụ thuộc vào sự thay đổi của tiêu thức nguyên nhân nói trên mà nó còn phụ thuộc vào các tiêu thức nguyên nhân khác, thống kê gọi mối liên hệ đó là mối liên hệ tương quan. Để phân tích và đánh giá mối liên hệ tương quan giữa các tiêu thức người ta thường sử dụng phương pháp tương quan, trên cơ sở quan sát hiện tượng số lớn. Phương pháp tương quan thường được vận dụng trong trường hợp trị số của tiêu thức nào đó bị thay đổi do ảnh hưởng của nhiều tiêu thức khác, trong đó mức độ ảnh hưởng của các tiêu thức có thể khác nhau. Tùy theo mục đích nghiên cứu có thể chọn riêng ra 1, 2 hoặc 3, 4 tiêu thức có ý nghĩa quyết định để xác định mối liên hệ tương quan giữa chúng với nhau. Trong thống kê thường phân biệt:

- *Tiêu thức nguyên nhân:* Là loại tiêu thức mà sự thay đổi trị số của nó là nguyên nhân làm ảnh hưởng tới sự thay đổi trị số của tiêu thức khác (thường ký hiệu x - là tiêu thức nguyên nhân);

- *Tiêu thức kết quả:* Là loại tiêu thức chịu ảnh hưởng tác động của các tiêu thức nguyên nhân (trong thống kê ký hiệu y - tiêu thức kết quả).

Ứng với mỗi giá trị (trị số) x_i của tiêu thức x ta có một trị số y_j của tiêu thức y . Số liệu ban đầu của hai tiêu thức này được trình bày dưới dạng bảng.

4.1.2. Nhiệm vụ của phương pháp hồi quy và tương quan

1. Khái niệm phương pháp hồi quy và tương quan

Hồi quy và tương quan là phương pháp toán học được vận dụng để phân tích mối liên hệ tương quan giữa các hiện tượng nghiên cứu gồm nhiều yếu tố, giữa các yếu tố có ảnh hưởng lẫn nhau, liên hệ tương quan với nhau.

2. Nhiệm vụ của phương pháp hồi quy và tương quan

Xác định phương trình hồi quy nhằm biểu diễn mối liên hệ tương quan dưới dạng một hàm số bao gồm:

- + Phân tích bản chất của hiện tượng;
- + Chọn dạng hàm số phù hợp với hiện tượng số lớn đã quan sát và bản chất của hiện tượng;
- + Tính toán các tham số của phương trình hồi quy;

Đánh giá mức độ chặt chẽ của mối liên hệ tương quan bằng cách tính hệ số tương quan và tỷ số tương quan.

4.1.3. Ý nghĩa của phân tích hồi quy và tương quan

Phương pháp phân tích hồi quy và tương quan là phương pháp thường được sử dụng trong thống kê để nghiên cứu mối liên hệ giữa các hiện tượng, như mối liên hệ giữa các yếu tố đầu vào của quá trình sản xuất với kết quả sản xuất, mối liên hệ giữa thu nhập và tiêu dùng, mối liên hệ giữa phát triển kinh tế và phát triển xã hội,...

Phương pháp phân tích hồi quy và tương quan còn được vận dụng trong một số phương pháp nghiên cứu thống kê khác như phân tích dãy số thời gian, dự báo thống kê...

4.2. Hồi quy tương quan tuyến tính giữa hai tiêu thức số lượng

Tương quan tuyến tính là tương quan biểu thị mối liên hệ giữa hai (hay nhiều) tiêu thức phù hợp với một phương trình đường thẳng nào đó. Trong trường hợp này các giá trị thực nghiệm của tiêu thức nguyên nhân (x) và tiêu thức kết quả (y) thường được trình bày dưới dạng bảng.

Ví dụ: Có tài liệu về số lao động và giá trị sản xuất của 10 doanh nghiệp công nghiệp như sau:

Bảng 4.1

Lao động	GO (tỉ đồng)
60	9,25
78	8,73
90	10,62
115	13,64
126	10,93

169	14,31
198	22,10
226	19,17
250	25,20
300	27,50

Trong mỗi liên hệ giữa số lao động và giá trị sản xuất thì số lượng lao động là tiêu thức nguyên nhân - ký hiệu là x , giá trị sản xuất là tiêu thức kết quả - ký hiệu là y .

Tài liệu cho thấy: Nhìn chung, cùng với sự tăng lên của số lượng lao động thì giá trị sản xuất cũng tăng lên, nhưng cũng có trường hợp không hẳn như vậy - như doanh nghiệp thứ hai so với doanh nghiệp thứ nhất: số lao động nhiều hơn nhưng giá trị sản xuất lại thấp hơn. Điều này chứng tỏ giữa số lượng lao động và giá trị sản xuất có mối liên hệ không hoàn toàn chặt chẽ, tức là liên hệ tương quan.

Giả sử giữa hai tiêu thức này có mối liên hệ tương quan tuyến tính theo một phương trình đường thẳng:

$$\widehat{y}_x = a + bx$$

Trong đó: x - giá trị thực nghiệm của tiêu thức nguyên nhân;

\widehat{y}_x - giá trị lý thuyết của tiêu thức kết quả tính theo phương trình hồi qui;

a, b - là các tham số qui định vị trí của phương trình hồi qui.

Quá trình xác định và đánh giá mối tương quan tuyến tính giữa hai tiêu thức số lượng được tiến hành theo trình tự sau:

Các hệ số a, b được xác định bằng phương pháp bình phương bé nhất:

$$\sum (y - \widehat{y}_x)^2 = \text{Min}$$

Từ đó ta có hệ phương trình sau:

$$\text{Hay: } \begin{cases} na + b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum x_i y_i \end{cases}$$

Hệ phương trình trên gọi là hệ phương trình chuẩn tắc xác định các tham số của phương trình hồi qui.

Giải hệ phương trình trên tính được a, b theo công thức sau:

$$b = \frac{\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\sum_i x_i^2 - \sum_i x_i \sum_i y_i} = \frac{\overline{x_i y_i} - \bar{x}_i \bar{y}_i}{\overline{x_i^2} - (\bar{x}_i)^2} = \frac{\overline{x_i y_i} - \bar{x}_i \bar{y}_i}{\delta_x^2}$$

$$a = \bar{y} - b\bar{x}$$

Thay các giá trị a, b vào phương trình hồi quy, ta sẽ có phương trình hồi qui biểu diễn mối liên hệ tương quan giữa hai giữa 2 tiêu thức số lượng x,y. Trong đó:

+ a là hệ số tự do, phản ánh ảnh hưởng của các tiêu thức khác đến tiêu thức kết quả y ngoài tiêu thức x

+ b là hệ số hồi quy, phản ánh mức độ ảnh hưởng của tiêu thức nguyên nhân x đến tiêu thức kết quả y.

Để tìm a, b cần tính $\sum x_i, \sum y_i, \sum x_i y_i, \sum x_i^2$ bằng cách lập bảng sau:

x	y	xy	x ²	y ²
60	9,25	555,00	3600	85,5625
78	8,73	680,94	6084	76,2129
90	10,62	955,80	8100	112,7844
115	13,64	1568,60	13225	186,0496
126	10,93	1377,18	15876	119,4649
169	14,31	2418,39	28561	204,7761
198	22,10	4375,80	39204	488,4100
226	19,17	4332,42	51076	367,4889
250	25,20	6300,00	62500	635,0400
300	27,50	8250,00	90000	756,2500
$\sum x = 1612$	$\sum y = 161,45$	$\sum xy = 30814,13$	$\sum x^2 = 318226$	$\sum y^2 = 3032,039$

Thay số liệu vào hệ phương trình trên:

$$\begin{cases} 10a + 1612b = 161,45 \\ 1612a + 318226b = 30814,13 \end{cases}$$

Giải hệ phương trình ta được:

$$a = 2,927, b = 0,082$$

Mô hình hồi quy tuyến tính phản ánh mối liên hệ giữa số lượng lao động và giá trị sản xuất là:

$$\widehat{y}_x = 2,927 + 0,082x$$

$a = 2,927$, nói lên các nguyên nhân khác, ngoài x , ảnh hưởng đến GO.

$b = 0,082$, nói lên khi thêm một lao động thì GO tăng bình quân 0,082 tỷ đồng.

Ví dụ trên đây nhằm trình bày phương pháp xây dựng mô hình hồi quy nên số lượng đơn vị được nghiên cứu không nhiều. Trong thực tế, số lượng được nghiên cứu có thể lên đến hàng trăm đơn vị, khi đó các chấm trên đồ thị sẽ rất nhiều và tạo thành một đám mây. Nhiều kinh nghiệm nghiên cứu cho thấy: Nếu đám mây có dạng hình elip hoặc hình bình hành thì có thể xây dựng mô hình hồi quy tuyến tính.

- *Hệ số tương quan:*

Hệ số tương quan là một số tương đối (biểu hiện bằng số lần) dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính. Hệ số này được tính từ các số liệu thực nghiệm của tiêu thức x và y . Căn cứ vào các giá trị tính được của nó có thể đánh giá được mối liên hệ tương quan tuyến tính có chặt chẽ hay không và chặt chẽ tới mức độ nào?

Hệ số tương quan có thể được tính theo công thức sau:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Biến đổi ta được:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Trong đó: x_i, y_i - là các trị số thực nghiệm của tiêu thức nguyên nhân và tiêu thức kết quả;

\bar{x}, \bar{y} - là số trung bình cộng của các trị số x_i, y_i

σ_x, σ_y - độ lệch tiêu chuẩn theo các tiêu thức x và y .

Từ ví dụ trên:

$$r = \frac{3081,413 - 161,2 \times 16,145}{\sqrt{5837,16 - 42,54}} = 0,961$$

Hoặc:

$$r = 0,082 \frac{\sqrt{5837,16}}{\sqrt{42,54}} = 0,961$$

Tính chất: Giá trị của hệ số tương quan r luôn nằm trong khoảng $-1 \leq r \leq 1$.

+ Khi $r = +1$ hoặc $r = -1$ thì liên hệ giữa tiêu thức x, y là liên hệ hàm số.

+ Khi $r = 0$ thì giữa hai tiêu thức x và y không có mối liên hệ tương quan.

+ Khi r càng gần giá trị ± 1 thì mối liên hệ giữa tiêu thức x và y càng chặt chẽ và ngược lại.

+ Khi $1 > r > 0$ giữa x và y có mối tương quan tỉ lệ thuận.

+ Khi $-1 < r < 0$ giữa x và y có mối liên hệ tương quan tỉ lệ nghịch.

Trong ví dụ trên, $r = 0,961$ cho thấy: mối liên hệ giữa số lượng lao động và giá trị sản xuất rất chặt chẽ và đây là mối quan hệ tỉ lệ thuận.

4.3. Hồi quy tương quan phi tuyến tính giữa hai tiêu thức số lượng

Tương quan phi tuyến là tương quan biểu thị mối quan hệ giữa các tiêu thức phù hợp với một đường cong nhất định (parabol, hypebol, hàm số mũ, loga...).

Khi giữa hai tiêu thức x và y tồn tại mối liên hệ tương quan phi tuyến thì phương trình đặc trưng của mối liên hệ tương quan này không phải là một đường thẳng mà là một dạng đường cong nào đó (parabol, hypebol, hàm số mũ, loga...). Cách thức chung để xác định dạng phương trình hồi qui dạng phi tuyến cũng tương tự như khi xác định mối liên hệ tương quan tuyến tính.

4.3.1 Các dạng phương trình hồi qui

1. Phương trình Parabol bậc 2

Phương trình Parabol bậc 2 thường được sử dụng khi các giá trị của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả tăng (hoặc giảm), việc tăng (hoặc giảm) đạt đến trị số cực đại (hoặc cực tiểu) rồi sau đó giảm (hoặc tăng). Phương trình hồi qui có dạng:

$$y_x = a + bx + cx^2$$

Trong đó: a, b, c là các tham số, được xác định bằng phương pháp bình phương nhỏ nhất, bằng cách giải hệ phương trình:

$$\begin{cases} na + b \sum x_i + c \sum x_i^2 = \sum y_i \\ a \sum x_i + b \sum x_i^2 + c \sum x_i^3 = \sum x_i y_i \\ a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 = \sum y_i x_i^2 \end{cases}$$

2. Phương trình Hypebol

Phương trình Hypebol được áp dụng trong trường hợp khi các trị số của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả có thể giảm và đến một giới hạn nào đó ($y_x = a$) thì hầu như không giảm. Phương trình hồi quy có dạng: $y_x = a + \frac{b}{x}$

Các tham số a, b được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na + b \sum \frac{1}{x_i} = \sum y_i \\ a \sum \frac{1}{x_i} + b \sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i} \end{cases}$$

3. Phương trình hàm mũ:

Phương trình hàm mũ được áp dụng trong trường hợp cùng với sự tăng lên của các trị số tiêu thức nguyên nhân thì các trị số của các tiêu thức kết quả thay đổi theo cấp số nhân, nghĩa là có tốc độ phát triển xấp xỉ nhau. Phương trình hồi quy có dạng: $y_x = a.b^x$

Các tham số a, b được xác định bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} n \lg a + \lg b \sum x_i = \sum \lg y_i \\ \lg a \sum x_i + \lg b \sum x_i^2 = \sum (\lg y_i) x_i \end{cases}$$

Ngoài 3 dạng phương trình phi tuyến ở trên, còn có nhiều dạng khác như Parabol bậc 3, lũy thừa, compec...

4.3.2. Các chỉ tiêu đánh giá mối liên hệ tương quan phi tuyến tính

Sau khi tìm được phương trình hồi quy ta đi đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến. Thông thường để đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến bằng bằng tỷ số tương quan.

Khi có mối liên hệ tương quan giữa tiêu thức nguyên nhân x và tiêu thức kết quả y thì có thể tính các loại phương sai sau đây:

+ Phương sai chung, phản ánh độ biến thiên của y do tất cả các tiêu thức nguyên nhân, trong đó có cả tiêu thức x :

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$$

+ Phương sai phản ánh độ biến thiên của y do ảnh hưởng riêng của tiêu thức x :

$$\sigma_{y_x}^2 = \frac{\sum (y_x - \bar{y})^2}{n}$$

+ Phương sai phản ánh độ biến thiên của y do ảnh hưởng của các tiêu thức nguyên nhân khác trừ tiêu thức x:

$$\sigma_{y(x)}^2 = \frac{\sum (y_i - y_x)^2}{n}$$

Mối quan hệ giữa 3 phương sai:

$$\sigma_y^2 = \sigma_{y_x}^2 + \sigma_{y(x)}^2$$

Như vậy ta thấy nếu tiêu thức nguyên nhân x ảnh hưởng đến tiêu thức kết quả y càng lớn thì tỷ trọng của phương sai $\sigma_{y_x}^2$, phản ánh độ biến thiên của tiêu thức y do ảnh hưởng của tiêu thức x chiếm trong phương sai chung σ_y^2 càng nhiều và ngược lại. Do đó tỷ số giữa phương sai $\sigma_{y_x}^2$ và phương sai chung σ_y^2 có thể dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan.

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{y(x)}^2}{\sigma_y^2}}$$

Hay:

$$\eta = \sqrt{1 - \frac{\sum (y_i - \bar{y}_x)^2}{\sum (y_i - \bar{y})^2}}$$

Trong đó: η - Tỷ số tương quan

y_i - là trị số thực nghiệm của tiêu thức kết quả;

y_x - là trị số lý thuyết của tiêu thức kết quả, tính theo phương trình hồi quy;

\bar{y} : là số trung bình số học của các trị số thực nghiệm.

Tính chất của tỷ số tương quan:

Tỷ số tương quan lấy giá trị trong khoảng $0 \leq \eta \leq 1$

+ Nếu $\eta = 0$ giữa x và y không có mối liên hệ tương quan;

+ Nếu $\eta = 1$ giữa x và y có liên hệ hàm số ;

+ Nếu η càng gần 1 mối liên hệ giữa x và y càng chặt chẽ và ngược lại.

Trong cùng một tiêu thức nghiên cứu x và y tỷ số tương quan bao giờ cũng lớn hơn giá trị tuyệt đối của hệ số tương quan tuyến tính. Nếu $\eta > |r|$ thì giữa x và y có mối liên hệ tương quan phi tuyến. Còn lại nếu $\eta = |r|$ thì giữa x và y có mối liên hệ tương quan tuyến tính.

4.4. Hồi quy tương quan tuyến tính bội

Trong thực tế một tiêu thức thường có nhiều mối liên hệ với các tiêu thức khác. Khi phân tích mối liên hệ giữa nhiều tiêu thức, cần căn cứ vào mục đích nghiên cứu để chọn các tiêu thức có liên hệ với nhau và chỉ chọn những tiêu thức có ý nghĩa nhất, ảnh hưởng lớn nhất đối với tiêu thức kết quả.

4.4.1. Mô hình tuyến tính bội

Để biểu diễn mối liên hệ tương quan giữa nhiều tiêu thức số lượng thường dùng phương trình tuyến tính. Phương trình tổng quát có dạng:

$$\overline{y_{x_1x_2...x_n}} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

Trong đó: x_1, x_2, \dots, x_n - là tiêu thức nguyên nhân;

y - là tiêu thức kết quả;

$a_0, a_1, a_2, \dots, a_n$ - là các tham số của phương trình hồi quy, thường được xác định bằng phương pháp bình phương nhỏ nhất.

Để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính giữa nhiều tiêu thức số lượng người ta thường sử dụng hệ số tương quan bội.

$$R_{y_{x_1x_2...x_n}} = \sqrt{1 - \frac{\sum (y - \overline{y_{x_1x_2...x_n}})^2}{\sum (y - \overline{y})^2}}$$

Trong đó: $R_{y_{x_1x_2...x_n}}$ - Hệ số tương quan bội, hệ số tương quan bội lấy giá trị trong khoảng $[0, 1]$. Nếu $R = 1$ thì có mối liên hệ hàm số;

$\overline{y_{x_1x_2...x_n}}$ - Giá trị lý thuyết tính theo phương trình hồi quy;

y - Giá trị thực nghiệm;

\overline{y} - Số trung bình số học của các giá trị thực nghiệm.

Ví dụ:

Có tài liệu về số lượng lao động, vốn đầu tư và giá trị sản xuất của 10 doanh nghiệp như sau :

Lao động	Vốn đầu tư (tỉ đồng)	GO (tỉ đồng)
60	1,8	9,25
78	1,1	8,73

90	1,9	10,62
115	2,5	13,64
126	1,3	10,93
169	2,6	14,31
198	5,1	22,10
226	4,2	19,17
250	7,5	25,20
300	6,1	27,50

Gọi :

x_1 : Số lượng lao động (người)

x_2 : Vốn đầu tư phát triển công nghiệp (Tỉ đồng)

y : Giá trị sản xuất (Tỉ đồng)

Giả sử có phương trình biểu diễn mối liên hệ tương quan tuyến tính giữa 3 tiêu thức, trong đó có 2 tiêu thức nguyên nhân (x_1 và x_2) và tiêu thức kết quả (y):

$$\bar{y}_{x_1x_2} = a_0 + a_1x_1 + a_2x_2$$

Các tham số a_0, a_1, a_2 được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum x_1 y \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum y x_2 \end{cases}$$

Căn cứ vào hệ phương trình để lập bảng tính toán sau :

x_1	x_2	Y	$y x_1$	$y x_2$	$x_1 x_2$	x_1^2	x_2^2
60	1,8	9,25	555,00	16,650	108,00	3600	3,24
78	1,1	8,73	680,94	9,603	85,80	6084	1,21

90	1,9	10,62	955,80	20,178	171,00	8100	3,61
115	2,5	13,64	1568,60	34,100	287,50	13225	6,25
126	1,3	10,93	1377,18	14,209	163,80	15876	1,69
169	2,6	14,31	2418,39	37,206	439,40	28561	6,76
198	5,1	22,10	4375,80	112,710	1009,80	39204	26,01
226	4,2	19,17	4332,42	80,514	949,20	51076	17,64
250	7,5	25,20	6300,00	189,000	1875,00	62500	56,25
300	6,1	27,50	8250,00	167,750	1830,00	90000	37,21
$\sum x_1$ = 1612	$\sum x_2$ = 31,4	$\sum y$ = 161,45	$\sum x_1 y =$ 30814,13	$\sum x_2 y =$ 681,92	$\sum x_1 x_2 =$ 6919,50	$\sum x_1^2 = 318226$	$\sum x_2^2 =$ 159,87

Thay số liệu vào hệ phương trình trên:

$$\begin{cases} 10a_0 + 1612a_1 + 34,1a_2 = 161,45 \\ 1612a_0 + 318226a_1 + 6919,50a_2 = 30814,13 \\ 34,1a_0 + 6919,50a_1 + 159,87a_2 = 681,92 \end{cases}$$

Giải hệ phương trình ta được :

$$a_0 = 3,775$$

$$a_1 = 0,042$$

$$a_2 = 1,646$$

Do đó mô hình hồi quy phản ánh mối liên hệ giữa vốn đầu tư phát triển công nghiệp, số lượng lao động với giá trị sản xuất của 10 doanh nghiệp này là :

$$\bar{y}_{x_1 x_2} = 3,775 + 0,042x_1 + 1,646x_2$$

Hệ số hồi quy chuẩn hóa : ký hiệu beta, được sử dụng đánh giá mức độ ảnh hưởng của từng tiêu thức nguyên nhân đối với tiêu thức kết quả và được tính theo công thức :

$$beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$$

Dấu của $beta_i$ là dấu của b_i phản ánh chiều hướng mối liên hệ là thuận hay nghịch giữa tiêu thức nguyên nhân với tiêu thức kết quả.

$|beta_i|$ phản ánh mức độ ảnh hưởng của từng tiêu thức nguyên nhân đối với tiêu thức kết quả.

Từ ví dụ trên:

$$beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,042 \times \frac{76,401}{6,522} = 0,492$$

$$beta_2 = b_{2i} \frac{\sigma_{x_2}}{\sigma_y} = 1,646 \times \frac{2,088}{6,522} = 0,527$$

Như vậy, vốn đầu tư phát triển công nghiệp và số lượng lao động đều quan hệ tỉ lệ thuận với giá trị sản xuất công nghiệp và ảnh hưởng của vốn đầu tư phát triển công nghiệp đối với giá trị sản xuất công nghiệp lớn hơn ảnh hưởng của số lượng lao động đối với giá trị sản xuất công nghiệp.

Hệ số tương quan bội và hệ số tương quan riêng phần: Để đánh giá trình độ chặt chẽ mối liên hệ tương quan sử dụng hệ số tương quan bội theo công thức:

$$R_{y_{x_1 x_2}} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}}$$

Trong đó

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x_1} \overline{y}}{\delta_{x_1} \delta_y}; \quad r_{yx_2} = \frac{\overline{x_2 y} - \overline{x_2} \overline{y}}{\delta_{x_2} \delta_y}; \quad r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \overline{x_2}}{\delta_{x_1} \delta_{x_2}}$$

Từ ví dụ trên:

$$R = \sqrt{\frac{0,961^2 + 0,965^2 - 2 \times 0,965 \times 0,892}{1 - 0,892^2}} = 0,99$$

Đánh giá trình độ chặt chẽ mối liên hệ tương quan riêng giữa từng tiêu thức nguyên nhân x_1, x_2 với tiêu thức kết quả y dùng hệ số tương quan riêng phần. Hệ số tương quan riêng phần được sử dụng để đánh giá mức độ chặt chẽ giữa một tiêu thức nguyên nhân nào đó với tiêu thức kết quả y trong khi các tiêu thức nguyên nhân khác không đổi.

Hệ số tương quan riêng phần giữa tiêu thức x_1 và tiêu thức y :

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 x_2}^2)}}$$

Hệ số tương quan riêng phần giữa x_2 và y :

$$r_{yx_2}(x_1) = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2)(1-r_{x_1x_2}^2)}}$$

Ta có:

$$r_{yx_1(x_2)} = \frac{0,961 - 0,965 \times 0,892}{\sqrt{(1 - 0,965^2)\sqrt{(1 - 0,892)^2}}} = 0,845$$

$$r_{yx_2(x_1)} = \frac{0,966 - 0,961 \times 0,892}{\sqrt{(1 - 0,961^2)\sqrt{(1 - 0,892)^2}}} = 0,862$$

4.4.2. Đa cộng tuyến

Khi xây dựng mô hình hồi quy giữa nhiều tiêu thức, về phương diện lý thuyết phải đảm bảo các tiêu thức nguyên nhân x_i không tương quan với nhau. Nếu giữa các tiêu thức nguyên nhân x_i có tương quan tuyến tính với nhau thì được gọi là hiện tượng đa cộng tuyến.

Hậu quả của đa cộng tuyến là làm cho việc ước lượng các hệ số của mô hình hồi quy sẽ không chính xác, ảnh hưởng đến việc suy rộng các kết quả tính toán.

Để khắc phục hiện tượng đa cộng tuyến, trong một số chương trình về thống kê có một số phương pháp xây dựng mô hình hồi quy sau đây:

- Phương pháp đưa vào dần: tiêu thức nguyên nhân đầu tiên được xem xét để đưa vào mô hình hồi quy là tiêu thức nguyên nhân có hệ số tương quan lớn nhất (về giá trị tuyệt đối) với tiêu thức kết quả. Để xem xét tiêu thức nguyên nhân này (và những tiêu thức nguyên nhân khác) có được đưa vào mô hình hồi quy hay không thì sử dụng tiêu chuẩn vào là thống kê F (mặc định $F=3,84$). Nếu tiêu thức nguyên nhân đầu tiên được xem xét để đưa vào mô hình hồi quy thỏa mãn tiêu chuẩn vào thì phương pháp đưa vào dần để tiếp tục, nếu không, không có tiêu thức nguyên nhân nào đưa vào mô hình hồi quy.

Khi tiêu thức nguyên nhân đầu đã được thỏa mãn tiêu chuẩn vào mô hình hồi quy thì tiêu thức nguyên nhân thứ hai được xem xét có thỏa mãn tiêu chuẩn vào hay không là tiêu thức nguyên nhân có hệ số tương quan riêng phần lớn nhất (về giá trị tuyệt đối) với tiêu thức kết quả. Nếu tiêu thức này thỏa mãn tiêu chuẩn sẽ được đưa vào mô hình hồi quy. Thủ tục này sẽ tiếp tục cho đến khi không còn tiêu thức nguyên nhân nào thỏa mãn tiêu chuẩn vào.

- Phương pháp loại trừ dần: Tất cả các tiêu thức nguyên nhân được đưa vào mô hình hồi quy. Sau đó loại trừ dần chúng bằng tiêu chuẩn loại trừ. Tiêu chuẩn loại trừ là giá trị F tối thiểu (được mặc định $F=2,71$) mà tiêu thức nguyên nhân phải đạt được để được ở lại trong mô hình hồi quy. Nếu các tiêu thức nguyên nhân có giá trị F nhỏ hơn giá trị F tối thiểu thì chúng bị loại khỏi mô hình hồi quy.

- Phương pháp chọn từng bước: Là sự kết hợp giữa hai phương pháp trên và là phương pháp thường được sử dụng.

Tiêu thức nguyên nhân đầu tiên được chọn để đưa vào mô hình hồi quy giống như phương pháp đưa vào dần, nếu nó không thỏa mãn tiêu chuẩn vào thì thủ tục này chấm dứt và không có tiêu thức nguyên nhân nào được chọn. Nếu nó thỏa mãn tiêu chuẩn vào thì tiêu thức nguyên nhân thứ hai được lựa chọn dựa vào hệ số tương quan riêng phần lớn nhất về giá trị tuyệt đối. Nếu tiêu thức nguyên nhân thứ hai thỏa mãn tiêu chuẩn vào thì nó cũng đi vào mô hình hồi quy.

Sau đó, dựa vào tiêu chuẩn ra để xem xét tiêu thức nguyên nhân thứ nhất có phải loại bỏ khỏi mô hình hồi quy hay không. Trong bước kế tiếp, các tiêu thức nguyên nhân không ở trong mô hình hồi quy được xem xét để đưa vào. Sau mỗi bước, các tiêu thức nguyên nhân ở mô hình hồi quy được xem xét để loại trừ ra cho đến khi không còn tiêu thức nguyên nhân nào thỏa mãn tiêu chuẩn ra thì kết thúc.

Các mô hình hồi quy được xây dựng theo các phương pháp trên có thể khác nhau. Tùy thuộc vào mục đích và nhiệm vụ nghiên cứu cụ thể để lựa chọn mô hình thích hợp.

TÀI LIỆU THAM KHẢO CHƯƠNG 4

1. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – Giáo trình **Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
2. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. TS. Hồ Sỹ Chi - **Thống kê doanh nghiệp**. NXB Tài chính, 2000
5. GS.TS. Phạm Ngọc Kiểm - **Giáo trình Thống kê doanh nghiệp**. NXB Lao động - Xã hội, 2012.
6. GS.TS. Phạm Ngọc Kiểm TS. Nguyễn Công Nhự - **Giáo trình Thống kê doanh nghiệp**. Nhà xuất bản Thống kê, 2007

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 4

1. Đặc điểm của liên hệ hàm số và liên hệ tương quan? Tại sao khi nghiên cứu các hiện tượng kinh tế xã hội lại thường gặp liên hệ tương quan.
2. Phân tích hồi quy và tương quan giải quyết những nhiệm vụ nghiên cứu gì?
3. Nêu những tính chất của r ?
4. Ý nghĩa của hệ số hồi quy chuẩn hóa?
5. Có tài liệu về tỉ lệ sinh đặc trưng theo tuổi của một địa phương như sau:

Tuổi	Tỉ lệ sinh
15-19	0,035
20-24	0,197
25-29	0,209
30-34	0,155

35-39	0,100
40-44	0,049
45-49	0,014

Yêu cầu:

- a. Xác định tiêu thức nguyên nhân và tiêu thức kết quả.
- b. Biểu hiện mối liên hệ bằng đồ thị.
- c. Xác định mô hình hồi quy phản ánh mối liên hệ giữa tiêu thức nguyên nhân và tiêu thức kết quả.
- d. Đánh giá mức độ chặt chẽ của mối liên hệ.